



Vetenskapsrådet



Sub-project, part of the SKOLFORSK-project

INTERNATIONAL COMPARISONS OF SCHOOL RESULTS

**A Systematic Review of Research on Large
Scale Assessments in Education**

INTERNATIONAL COMPARISONS OF SCHOOL RESULTS – A SYSTEMATIC REVIEW OF RESEARCH ON LARGE SCALE ASSESSMENTS IN EDUCATION

Swedish Research Council
Box 1035
SE-101 38 Stockholm, SWEDEN

© Swedish Research Council
ISBN 978-91-7307-288-5

**INTERNATIONAL COMPARISONS OF SCHOOL RESULTS
– A SYSTEMATIC REVIEW OF RESEARCH ON LARGE SCALE
ASSESSMENTS IN EDUCATION**

Sverker Lindblad, Daniel Pettersson, Thomas S. Popkewitz

A Report from the Educational Research project SKOLFORSK

The Swedish Research Council carried out a project in 2014 in Educational Research, called SKOLFORSK. The work was commissioned by the government to survey Swedish and international research that is relevant to education. The aim was to create a platform of knowledge for the newly-formed Swedish Centre for Educational Research (Skolforskningsinstitutet).

The conclusions of this report are those of the authors. The summary report by the Swedish Research Council, "Research and Education in collaboration" with a description of the project and the issues, results and recommendations presented in the sub-projects, as well as the other reports, can be downloaded from the Swedish Research Council's website vr.se.

Authors

Sverker Lindblad

Daniel Pettersson

Thomas S. Popkewitz

THE AUTHORS' PREFACE

This is a report to the Swedish Research Council and the new Swedish Centre for Educational Research from a systematic research review on international comparisons of school performances. The text is written to function as a technical report to the Centre, providing detailed information on how the review was carried out as well as our review of how we did the review. If given opportunities it is our ambition to present the major contents of the report as a book also addressing different stake-holders in such reviews, such as teachers and students as well as policy-makers and interested citizens. We will come back on that issue later.

Our research is a journey. We had the task to systematically review international comparisons of school performances. We had previously been doing research reviews and shared the ambition in such reviewing to be transparent and systematic – for instance trying to avoid the fallacy of “*cherry-picking*” where reviewers refer to what they happen to know or select studies which supported their position. However, the dominant discourse on systematic research also contained a few characteristics that seemed to be contra-productive in relation to the field of study we were dealing with.

In an early phase we realized that so called aggregated studies meta-analyzing research results was actually not a possible or meaningful thing to do, give the field of study that we were dealing with. Thus, we turned into a more conceptual approach in reviewing; focusing on the ways different studies were framing their research problematic as well as results. We spent considerable time to classify, filter, code and map such framework and could identify the heterogeneity of study objects in international comparative research on school results. In doing so we realized that there was a need to further analyze the conceptual frameworks of these studies – what they tried to explain, what inquires that was carried out in the studies and what they conceived of as results. Here, we did not only find heterogeneity but also could identify some dominant themes in the research and how this in turn was used as ways to talk about education and what was needed to be done with educational problems in policy as well as research. Furthermore, we realized that there was a common style – though displayed in different ways – of statistical reasoning in the studies analyzed and that this style is important to capture when translating research into practice.

Given the large amount of publications that we found – more than 11000 academic texts that explicitly was referencing to education and various ILSA (International Large-Scale Assessments) in their abstracts – we decided to describe this terrain and focus on the parts that were published in peer reviewed scientific journals, while other texts which had not passed through the critical eyes of scientific reviewing were excluded. An implication of this was that texts that were central in policy discourses – such as e.g. the so called McKinsey-reports and OECD national report that draw on ILSA would not be part of our data.

During this journey we decided that there was a need to describe the research field in a systematic and transparent way that is also reflected characteristics in this discursive field intersecting with policy discourses and ambitions to improve or change education. In doing this, we had to reconsider the omission as the public reports of McKinsey and OECD are central ingredient of the field and thus we made this a part of our conceptual analyses. We considered this as necessary in order to understand the constraints and opportunities that the current field of research is offering as systems of reasoning in defining educational terrains and changes to be achieved in these terrains. In that it can be understood as a way to deal with the politics of knowledge in educational systems.

It is our ambition in this systematic review to present analyses of how the existing corpus of research constructs its objects and arguments; what are the rules and standards that are being used to identify and measure school performances, differing ways to capture effectiveness and efficiency as well as ways of ordering and classifying students and their schools and educational systems. It is our hope that this reviewing will be clarifying in order to capture results and impact of international comparisons of school results.

Many thanks to the colleagues who commented our presentations at research council conferences! Special thanks to Corinne Erfurth, who edited our manuscript in a most helpful and engaged way!

THE COUNCIL'S PREFACE

On 21 November 2013, the government gave the Swedish Research Council the commission (U2013/6845/S) of arranging validated surveys of Swedish and international research results that are relevant to the educational system. The surveys would be based on issues that are relevant to, and developed in consultation with, staff at schools and pre-schools. The purpose of the surveys was to form a basis for systematic summaries of research results that are relevant to activities in schools and pre-schools that the Swedish Centre for Educational Research would then be commissioned to carry out. The commission was formulated after the Secretary General of the Committee for Educational Sciences at the Swedish Research Council made a proposal for a number of projects which would produce documentation for the Swedish Centre for Educational Research over a period of one year.

Despite the short time available, the commission from the government, under the working name SKOLFORSK (Educational Research), resulted in 16 sub-projects in which around 40 researchers from 15 different universities in Sweden, Norway and the United States have participated. For each project, one well-qualified researcher with expert knowledge in the relevant area was in charge. The sub-projects, which were all carried out in 2014, varied in length from 4 to 11 months. The short studies are aimed at making the initial work of the new authority easier. This work relates to processes and models for creating new knowledge and favourable conditions for the use of research-based knowledge in schools. The long projects are examples of different types of systematic summaries of research results. They illustrate various models and methods of how research on learning in school can be systematised and made visible.

The Secretary General of the Swedish Research Council, Professor Eva Björck, and the project manager, Ph.D. Cristina Robertson, have been responsible for the project. SKOLFORSK has had a consultative group with different players who are working to promote practical research and dissemination of research. The project has had close contact with the group which planned the Swedish Centre for Educational Research.

A warm word of thanks is sent to all the researchers who made it possible to carry out this project at such short notice. You have enriched the school system and the Swedish Centre for Educational Research with a sound base for continued work with the school's scientific foundation and building up practical educational research in Sweden for the benefit of pre-schools, schools and teacher training.

We wish the Swedish Centre for Educational Research much success in its future work!

Petter Aasen, Chairman, Committee for Educational Sciences

Eva Björck, Secretary General for Educational Sciences at the Swedish Research Council

TABLE OF CONTENTS

| | |
|--|----|
| THE AUTHORS' PREFACE | 3 |
| THE COUNCIL'S PREFACE | 5 |
| THE AUTHORS | 10 |
| SVENSK SAMMANFATTNING | 11 |
| Översikt | 11 |
| Inledning och inramning | 11 |
| Teoretisk ansats | 11 |
| Studiens frågeställningar | 12 |
| Studiens design och metod | 12 |
| Resultat | 13 |
| Slutsatser | 14 |
| Noteringar om forskarsamhället med dess publiceringsmönster och tankestilar:..... | 14 |
| Översättningar | 15 |
| Kartläggningsmodellens starka och svaga sidor, hinder och möjligheter | 15 |
| Särskilt viktigt vid inrättande av ett Skolforskningsinstitutet | 16 |
| Rekommendationer till Skolforskningsinstitutet..... | 16 |
| Litteratur | 16 |
| THE RESEARCH REVIEW PURPOSE AND PROBLEMATICS | 18 |
| Background | 18 |
| The field of study and style of reasoning..... | 20 |
| Purpose and research questions..... | 21 |
| The research review problematics | 22 |
| Explanandum and explanans..... | 22 |
| Practical reason..... | 23 |
| Review problematics | 23 |
| ON THE DESIGN OF THE RESEARCH REVIEW | 24 |
| On systematic research reviews | 24 |
| Evidence and validity of systematic research reviews | 25 |
| Approaches to systematic research reviews | 26 |
| Design procedures for searching and filtering publications..... | 27 |
| The core problematic..... | 27 |
| The field problematic | 29 |
| The translation problematic | 29 |
| Search tool: Discovery systems in systematic reviews | 29 |
| Difference between "discovery" and "search" | 29 |
| Search with discovery systems | 30 |
| New kids on the block..... | 30 |
| A HISTORY OF COMPARATIVE EDUCATION AND PROGRAMS FOR INTERNATIONAL LARGE-SCALE ASSESSMENTS..... | 33 |
| A short history of comparative education studies..... | 33 |
| Comparing education systems and organizations | 33 |

| | |
|--|----|
| Numbers and comparisons – style of reasoning | 34 |
| Methodologies for comparisons: Two trajectories | 36 |
| Spaces for comparison and data: From comparing to collaboration and learning | 38 |
| A history of large-scale assessments | 39 |
| International Assessments | 39 |
| Regional Assessments | 44 |
| National Assessments | 47 |
| From research to policy and back: assessing impact | 48 |
| Growth of ILSA as a measure of impact | 49 |
| Discourse as a measure of impact | 51 |
| Changes in educational policy as a measure of impact | 51 |
| Curriculum changes as a measure of impact | 52 |
| Changes in teaching as a measure of impact | 52 |
| Capacity building and research endeavors as a measure of impact | 52 |
| Global and donor responses as a measure of impact | 54 |
| Research generating as a measure of impact | 54 |
| Concluding comment | 55 |
| SELECTED RESEARCH PROGRAMS | 56 |
| PISA | 56 |
| TIMSS | 56 |
| Civic studies – ICCS and CIVED | 58 |
| A REVIEW OF PISA PUBLICATIONS | 59 |
| Field review and filtration process | 59 |
| Inclusion and exclusion of articles | 59 |
| Observations on filtering processes | 61 |
| Screening and coding selected publications | 62 |
| Observations on the screening and coding processes | 63 |
| Mapping explanandum and explanans in PISA studies | 63 |
| Explanandum | 64 |
| Explanans | 64 |
| An example of mapping practices | 65 |
| Results of mapping PISA publications | 66 |
| Observations on the mapping process | 67 |
| Synthesizing | 67 |
| Efficiency problematic research | 68 |
| Equity inquiries | 70 |
| Student direction problematics | 75 |
| Observations on the synthesizing processes of equity studies | 77 |
| Reviewing the PISA research review | 78 |
| A REVIEW OF THE TIMSS STUDY | 81 |
| Field review and filtration process | 81 |
| Inclusion and exclusion of articles | 81 |
| Observations on filtering processes | 81 |
| Screening and coding selected TIMSS publications | 84 |

| | |
|---|-----|
| Mapping explanandum and explanans in TIMSS studies | 85 |
| Explanandum and explanans | 85 |
| Mapping TIMSS publications | 87 |
| Synthesizing | 88 |
| Efficiency inquiries | 89 |
| Equity inquiries | 95 |
| Randomized inquiries | 100 |
| On TIMSS research compared to PISA research | 103 |
| REVIEW OF THE CIVED/ICCS STUDIES | 104 |
| Field review and filtration process | 104 |
| Inclusion and exclusion of articles | 104 |
| Screening and coding selected publications | 107 |
| Mapping explanandum and explanans in CIVED/ICCS studies | 109 |
| Synthesizing | 110 |
| Summarizing CIVED/ICCS research | 115 |
| COMPARISONS OF REVIEWS OF PISA, TIMSS AND CIVED | 116 |
| The filtration process of articles | 116 |
| The coding of articles | 118 |
| The impact of PISA, TIMSS and CIVED/ICCS on academic journals | 120 |
| In which journals are the explicit articles published? | 120 |
| What kinds of articles are most commonly disseminated? | 121 |
| Who are the researchers? | 126 |
| In which journals is the ILSA research disseminated? | 127 |
| In conclusion | 128 |
| THE GREYZONE: AN IN-BETWEEN SPACE OF INTERVENTION MODELS IN MCKINSEY AND OECD | 130 |
| Introduction | 130 |
| The ILSA grey-zone | 130 |
| Educational change and managing complexity: Contributions. | 132 |
| A Conceptual Approach: Models of systems as fabrications | 133 |
| Systems | 134 |
| Schools as organisms (equilibrium/disequilibrium) | 135 |
| Self-referential/Self-authentication | 137 |
| Developing equivalents as universals to actualize | 139 |
| The search for certainty | 141 |
| Illusion of democracy and equity | 141 |
| Inclusion/Exclusion | 142 |
| Cherry picking | 143 |
| Conclusions | 144 |
| CONCLUSIONS AND DISCUSSION | 146 |
| Introduction | 146 |
| Reviewing the reviews | 146 |
| Review results – arguments and conclusions | 146 |
| The research field of ILSA – research communication and style of reasoning | 147 |

| | |
|--|-----|
| Analyzing translations of ILSA research..... | 148 |
| Considering the School Research Institute | 149 |
| A final point on the research review | 150 |
| LITERATURE | 152 |
| MAPPED ARTICLES | 158 |
| APPENDIX I | 166 |
| Classification of PISA, TIMSS and CIVED/ICCS articles..... | 166 |
| APPENDIX II: MAJOR PUBLICATIONS FROM ILSA..... | 193 |
| IEA Publications | 193 |
| OECD Publications (PISA) | 200 |
| APPENDIX III: SHEET FOR CODING ARTICLES | 209 |

THE AUTHORS

Sverker Lindblad is professor in Education Sciences, University of Gothenburg. He has coordinated international studies such as the European Commission funded research programs on *Education Governance and Social Inclusion/Exclusion* (EGSIE), *Professional Knowledge in Restructuring Welfare States* (PROFKNOW), and the RJ funded *Swedish School Culture in International Perspective* (KULT) which is the Swedish part in an international consortium doing micro-ethnographic studies of classroom interaction from the learners perspective (the LPS project). Relevant publications for this project are e.g. *Från förskola till högskola i olika länder* (med Erik Wallin, 1983), *Curriculum Codes and International Statistics* (2014), *Educational Statistics and Social Integration and Exclusion* (with Thomas, S. Popkewitz, 2001).

Daniel Pettersson, University of Gävle is a Senior Lecturer in Education Sciences. He specializes in the history, dissemination and use of international large-scale assessments, in policy narratives and in the history of education. He is currently a member of the research group STEP, *Studies in Educational Policy and Educational Philosophy* at Uppsala University and member of the board and Chief Editor of the Nordic Journal of Studies in Educational Policy. Relevant publications for this project are e.g. *The Development of the IEA: The Rise of Large-scale Testing* (2014), *Three Narratives: National Interpretations of PISA* (2014) and *European Educational Transfer and Curriculum displacement* (2014).

Thomas S. Popkewitz, University of Wisconsin-Madison is a curriculum theorist and professor from the United States of America, on the faculty at the University of Wisconsin–Madison School of Education. His studies are concerned with the knowledge or systems of reason that govern educational policy and research related to pedagogy and teacher education. His research includes histories of the present, ethnographic and comparative studies of national educational reforms in Asia, Europe, Latin America, Southern Africa, and the US. His book *Cosmopolitanism and the Age of School Reform* (2008) explore the systems of reason in pedagogy through historically examining the changing images and narratives of Enlightenment concerns with cosmopolitanism. He has written or edited approximately 30 books and 200 articles in journals and book chapters. Two of his books (*Paradigms and Ideology in Educational Research and A Political Sociology of Educational Reform*) have won awards for their contribution to educational studies.

SVENSK SAMMANFATTNING

Översikt

Avsikten med denna systematiska forskningsöversikt är att beskriva och analysera forskning om internationella jämförelser av skolresultat genom storskaliga studier (International Large Scale Assessments, ILSA). Hur är kunskapsläget och vad kännetecknar forskningsfältet? Vilken relevans har forskningen? Vad betyder systematiska forskningsöversikter för kunskapsutveckling och expertis inom utbildningsområdet?

Vår genomgång inriktades på OECDs forskningsprogram PISA (Programme for Individual Study Assessment), och IEAs (International Association for the Evaluation of Educational Achievement) två program TIMSS (Trends in International Mathematics and Science Study) och CIVED/ICCS (International Civic and Citizenship Education Study) till vilka sammanlagt närmare 9000 publikationer kunde knytas.

I olika steg filterades publikationerna så att vi slutligen erhöll studier som kunde klassificeras som primärforskning av internationell komparativt slag i ”peer reviewed” vetenskapliga tidskrifter. Dessa studier kodades med fokus på argument och belegg. Det kartlagda forskningsfältet syntetiserades sedan inom avgränsade områden. Vi identifierade vissa sätt att resonera inom forskningsfältet - som resonemangs-stilar där vissa slutledningar är giltiga men inte andra.

Möjligheter och begränsningar att översätta den aktuella forskningen till praktisk verksamhet undersöktes dels med avseende på vad som skrevs i de aktuella publikationerna, dels genom specialundersökning av den betydelsefulla gråzon där PISA-resultat översattes till ranking av utbildningsystem och strategier för förbättringar av dessa system.

Slutsatser drogs om hur den valda modellen för forskningsgenomgångar fungerade. Vidare diskuterades premisser och konsekvenser av systematiska forskningsgenomgångar och hur de kan utnyttjas.

Inledning och inramning

Under det senaste decenniet har internationella jämförelser av skolresultat pekats på en svensk skola i utförsbacke. Skolverket har gjort nyanserade analyser av denna utveckling och visar på tre viktiga resultat:

- Kunskapsrenden för svensk grundskola är i ett internationellt perspektiv tydligt negativ inom matematik, naturkunskap och läsning.
- Likvärdigheten i svensk skola har försämrats sedan början av 2000-talet.
- Resultaten är i stort bättre för yngre elever i grundskolan.

Dessa storskaliga undersökningar av hur olika länder presterar på kunskapsmätningar har fått rejält genomslag i massmedia och i utbildningspolitiska diskussioner i Sverige liksom i många andra länder under senare år. Samtidigt har internationella jämförelser av utbildningsresultat har en lång historia. Vi går igenom denna utveckling med olika aktörer över tid, internationellt, regionalt och nationellt, och hur fältet expanderat över tid.

Teoretisk ansats

Vi arbetar med en speciell systematik för att kartlägga och förstå ett forskningsfält, vilken kunskap som genererats inom detta fält och hur den kan nyttiggöras inom utbildningssystemet. Detta innebär att vi inte bedriver ILSA-forskning som sådan, utan gör analyser av kunskapsproduktion, forskningsfält och hur översättningar till verksamheter inom utbildningsväsendet utformas.

För att kunna bedöma begränsningar och möjligheter inom detta forskningsfält är det rimligt att förstå vad som är objekt för forskningen och vad som i detta sammanhang räknas som belegg och objektiv kunskap. Därför utvecklade vi en ansats om resonemangs- eller slutledningsstilar inom vetenskaplig verksamhet (Hacking, 1992) – vad som studeras och vad som gäller för rimliga slutledningar i kontext av en sådan *tankestil* – och på motsvarande sätt inom utbildning (Popkewitz, 2014). Resonemang och slutledningar analyseras som

en fråga om relationer mellan ”vad som ska förklaras” (explanandum) och ”de undersökningar som sägs kunna förklara detta” (explanans). På så sätt beskriver vi hur argument formuleras och vilka slutsatser som kan dras. Frågor om översättning från forskningsresultat till utbildning ser i perspektiv av praktiskt förnuft (von Wright, 1983).

Studiens frågeställningar

Våra studier genomförs för att få svar på tre relaterade frågeställningar rörande storskaliga internationella jämförelser av skolresultat:

- A. Vilken omfattning och inriktning har primärforskningen inom forskningsfältet och vilka resultat har uppnåtts?
- B. Hur översätts och hanteras resultat av ILSA-forskning till påståenden av relevans för politiska och professionella bedömningar och beslut inom utbildningsområdet?
- C. Hur gör man en systematisk forskningsgenomgång inom det aktuella fältet – efter vilka principer och modeller är det rimligt att arbeta och med vilka konsekvenser?

Till detta kommer en fjärde frågeställning om innebörden av systematiska forskningsgenomgångar för kunskapsorganisering:

- D. Vad betyder systematiska forskningsgenomgångar för kunskapsbidrag och deras giltighet inom utbildningsområdet?

Studiens design och metod

Utifrån ovanstående kunskapsbehov och frågeställningar görs en systematisk genomgång enligt erkända principer och med hänsyn tagen till det aktuella mycket stora fältet av forskning som ILSA omfattar. Hur studien genomförs är en viktig del av vår undersökning och därför presenteras upplägget utförligt.

Med ambitionen att reflektera över innebörden av systematiska forskningsgenomgångar – vad de vilar på för grund och vilken kunskapspolitisk betydelse de har – försökte vi genomföra våra studier på en principiell nivå vid val av ansats och design (se Gough et al, 2013, Pröitz, 2015) och med en transparent systematik. Vår undersökning ägde rum i flera steg:

- 1) *Inriktning och avgränsning*: Vilka definitioner görs och vilka kriterier används för att inkludera och exkludera publikationer:
 - a) Identifiering av ILSA publikationer
 - b) Endast publikationer i peer-review journals inkluderas.
 - c) Enbart publikationer med resultat från empiriska studier.
- 2) *Sökning*: Identifieringen av publikationer genomfördes:
 - a) med hjälp av sökmaskinen Discovery och givna sökord
 - b) kompletterat med Google Scholar, ERIC och Swepub (för att identifiera relevanta publikationer som exkluderats under a)
- 3) *Översikt och organisering av publikationer*:
 - a) Kategorisering av publikationer.
 - b) Utifrån citeringar analysera ILSA som forskningsfält
 - c) Identifiering av relevanta fokus för forskningen
- 4) *Kodning*:
 - a) Utveckling av kodschema
 - b) kodning och sammanställning

- 5) *Kartläggning:*
 - a) Bestämning av vad som har publicerats och vad studierna för fram
 - b) Genomgång av argument över explanandum och explanans, samt slutsatser av praktisk relevans i studierna.
- 6) *Bedömning:*
 - a) Granskning av publikationer och deras kunskapsbidrag.
 - b) Granskning av forskningsgenomgången.
- 7) *Syntetisering:*
 - a) Vilka slutsatser kan dras av den genomgångna forskningen relativt explanandum och explanans?
 - b) Vilka iakttagelser har gjorts om forskningsfältet och om forskningsgenomgången?
- 8) *Kommunikation och granskning:*
 - a) Eftersom vi rör oss inom ett internationellt fält har vi skrivit på engelska och i arbetet kommunicerat med internationell expertis som fått granska vårt upplägg och våra analyser.
 - b) Resultaten av vårt arbete kommer att redovisas för och diskuteras med olika aktörer i fältet, som med lärarnas organisationer och utformas för att tas upp i lärarutbildningarna.

Resultat

Totalt gav sökningen en *omfattning* om 8744 träffar på publikationer som byggde på de aktuella forskningsprogrammen PISA, TIMSS och CIVED. Av dessa var 4786 publicerade i peer-reviewed vetenskapliga tidskrifter. Vi filtrerade bort irrelevanta tidskrifter, dubletter av artiklar, redaktionella texter m m och fick på så sätt kvar 353 artiklar. Eftersom vi fokuserade på empirisk primärforskning exkluderades ytterligare artiklar av metodologisk och policy-analytisk art. Av de återstående 289 artiklarna visade det sig att 154 inte gjorde internationella jämförelser. Efter denna *filtreringsprocess* återstod 135 artiklar som svarade upp mot de kriterier som vi ställt upp på förhand.

Genom denna filtrering med fokus på empirisk primärforskning publicerad i kollegialt granskade vetenskapliga tidskrifter exkluderades stora delar av publikationer inom ILSA. Exempelvis ingår inte OECDs egna publikationer eller de rapporter som konsultföretaget McKinsey publicerat om utbildningssystem och som befinner sig en viktig gråzon mellan forskning och politik. Vidare kunde forskning rörande internationella jämförelser av skolresultat publicerad på svenska inte identifieras i de sökningar som gjordes. Med anledning av detta gjordes kompletterande undersökningar där ytterligare ett mindre antal publikationer identifierades.

Över hälften av artiklarna skrevs av forskare med *hemvist* i USA, UK, Kanada och Australien och ytterligare en femtedel av forskare från nordvästra Europa. Publikationskanalerna varierar mellan forskningsprogram liksom förekomsten av citeringar. I snitt citeras de artiklar vi identifierat 6,8 gånger enligt våra analyser.

Kodning av artiklarna visade att merparten av forskningen gjorde sina undersökningar inom ramen för en och samma speciella databas samt att de var av explorativ art, medan mindre än en tredjedel testade mer specifika hypoteser. Det stora flertalet av undersökningarna fokuserade på elevprestationer medan mindre andelar analyserade egenskaper hos skolor och nationella skolväsenden. Det var ofta individer och deras egenskaper som stod i fokus, men undersökningar av organisering och utbildningssystem var inte ovanliga. Oftast förde man fram att undersökningarnas resultat var av utbildningspolitisk eller professionell relevans. Endast ett fåtal artiklar lyfte fram den vetenskapliga relevansen av den genomförda studien.

Kartläggningen av den undersökta internationellt jämförande forskningen analyserade resultat av prov i olika ämnen där publikationerna studerade kunskaper eller kompetenser i olika ämnen ofta i kombination med varandra. Kunskaper i matematik var oftast undersökt, följt av naturkunskap och läsning, och därefter kunskaper i samhällskunskap. Till bilden hör att ett mycket stort antal undersökningar behandlade andra frågor, som attityder till olika ämnen eller framtidsförväntningar, eller frågor om politisk participation och etnicitet.

Synteser genomfördes på basis av de argument som fördes fram för att förstå eller förklara de förhållanden som är för handen i skolväsendet och/eller de provresultat som erhållits. Tre grundläggande kategorier av problemområden kunde identifieras:

- 1) Prestationsklyftor: bygger på frågor om jämlikhet, jämställdhet eller likvärdighet i utbildningsresultat för olika kategorier av elever – ofta i termer av kön, socio-ekonomisk bakgrund, eller etnicitet. Analyserna knyts ofta till skillnader i att organisera och differentiera utbildning, exempelvis när uppdelning av elever sker i olika program eller relativt utformning av skolval. Frågor om social segregering är vanliga.
- 2) Framgångsfaktorer: Vilka utbildningssystem eller sätt att organisera skolor och undervisning ger bättre resultat frågar man sig här. Analyserna bygger mestadels på provresultat i olika ämnen i kombination med variabler som man menar vara av betydelse.
- 3) Kunskapskaraktistika: här undersöks elevernas kunskaper eller kompetenser på mer detaljerad nivå för att komma åt kunskapsorganisering och interaktion mellan kunskapsutveckling och kontext.

En viktig uppgift för systematiska forskningsöversikter är att dra slutsatser av praktisk art utifrån den samlade relevanta forskningen. Nedan följer exempel på översättningar till möjliga beslut art utifrån de synteser som vi gjort:

- I stort finns en samsyn inom forskningen att sen differentiering i olika skolformer eller program minskar riskerna för social segregering och socialt och kulturellt bestämda prestationsklyftor.
- Några väl utvecklade analyser kunde visa på komplexiteten hos utbildningssystem relativt prestationsklyftor. Andra studier visade hur valfrihet i olika kunde knytas till segregering i utbildning och samhälle.
- Insatser på sekundärskolenivå (högstadiet eller gymnasiet) har mindre betydelse för att förbättra positionen relativt andra länder. Insatserna bör istället göras så tidigt som möjligt i skolsystemet.

Noteras bör att forskningsfrågorna ofta färgas av de sammanhang – nationella kontexter – som forskarna verkar inom.

Slutsatser

Vi rörde vi oss i ett relativt avgränsat forskningsfält vad gäller kunskapsanspråk, datamaterial och analysmetoder. Däremot fanns det stora skillnader i såväl studieobjekt som kunskapsobjekt. Att ställa samman forskningsresultat när dessa skillnader är för handen vore att blanda och homogenisera distinkta olikheter. Vi valde därför att identifiera och särskilja olika slag av *explanandum* (vad som ska förklaras) och *explanans* (hur det kan förklaras) inom fältet för att kunna kartlägga och syntetisera forskningen. Behovet av sådana distinktioner vad gäller såväl kunskapsobjekt som studieobjekt gäller förmodligen för de flesta forskningsgenomgångar inom utbildningsområdet.

Noteringar om forskarsamhället med dess publiceringsmönster och tankestilar:

Kommunikationen inom forskarsamhället förefaller vara begränsad: Av det totala antalet publikationer är det endast en liten del (mindre än två procent) som utgörs av internationellt komparativ primärforskning publicerade i vetenskapliga tidskrifter. Även bland dessa är det en mindre andel som explicit riktar sig till forskarsamhället. Publiceringsmönstret är något spretigt vilket pekar i riktning mot att ett mer tydligt forskningsobjekt inte formulerats – kanske med undantag av forskning om prestationsklyftor och orsakerna till dessa.

Däremot finns en etablerad stil att hantera vad som är giltiga påståenden relativt vilka statistiska analyser som gjorts och styrkan mellan olika variabler. Mot bakgrund av detta blir det inte längre överraskande att endast en liten andel av publikationerna redovisar resultat av hypotesprövningar.

Den *tankestil* som vi menar etablerats inom ILSA bygger på att påståenden av avkontextualiserad och generaliserbar art som visar på styrkan mellan olika variabler eller distinktioner inom taxonomier – då ofta genom avancerade statistiska analyser – är de som inom forskarsamhället bedöms vara valida och hållbara. Vissa undersökningar förefaller vara mer utvecklade relativt en sådan tankestil.

Översättningar

Den kunskap som genereras inom ILSA produceras och bedöms inom en tankestil med bestämda objekt och vad som inom denna tankestil räknas som belägg. Detta får konsekvenser vad gäller möjligheterna till översättning in i utbildningsområdet där frågor om giltighet och *kontextualisering* är nödvändiga att hantera. Vår identifiering och genomgång av en gråzon med översättningar från forskningsresultat till strategiska direktiv för förbättring av utbildningssystem visade på rejäla på behov av konkretisering och kontextualisering. Våra analyser visade på problem för aktörer inom gråzonen att hantera helheten i problematiken. Det framstår som viktigt att klargöra den tankestil som ILSA-forskning vilar i och de påståenden som hävdas som giltiga inom denna tankestil för att översättningarna ska ha en innehållslig giltighet.

Med detta vill vi också ha sagt att det finns behov av kritiska analyser av hur ILSA översätts, används och utnyttjas i utbildningspolitiska och kunskapspolitiska sammanhang. Motsvarande slutsats gäller förstå även andra tankestilar inom utbildningsforskning som fallstudier – och även systematiska forskningsgenomgångar.

Kartläggningsmodellens starka och svaga sidor, hinder och möjligheter

Modellen med dess olika kriterier för urval och kategoriseringar av publikationer fungerade väl i termer av systematik och transparens. Vi kunde identifiera ett stort antal publikationer som filterades i olika steg i syfte att få fram en uppsättning av empirisk komparativ primärforskning publicerad i peer-reviewade tidskrifter. Detta filteringsarbete var mycket tidskrävande och krävde omdömesgillt hanterande även om de valda kriterierna fungerade väl.

Vad gäller filteringsprocessen bör några mer olyckliga utfall lyftas fram:

- Åtskilliga publikationer av visad betydelse inom utbildningspolitiska diskurser filterades bort.
- Vidare exkluderades rapporter och andra texter som inte återfanns inom vetenskapliga kollegialt granskade publikationer.
- Genom utnyttjandet av internationell sökmaskin sorterades svenskspråkig forskning i stort bort.

Vi noterade dessa konsekvenser. Det var en fördel att de blev synliga och kan bli föremål för diskussion och åtgärder. Vi kompletterade därför sökprocessen med fokus på svenska och svenskspråkiga publikationer. Utfallet blev dock ganska magert, vilket kan vara värt att notera av tre skäl. (a) Sökprocessen fungerar mindre väl och behöver revideras. (b) Svenska forskare behöver göra sig mer synliga för litteratursökningar. (c) Samtidigt visar vårt fall att sökprocessen förefaller ha en bias där stora delar av världens forskning troligen inte inkluderas och utnyttjas.

Vad gäller diskursivt betydelsefulla publikationer som filterades bort kan detta ses som en viktig gråzon mellan forskning och politik värd att uppmärksamma och analysera. Vidare sker exempelvis inom forskning som studerar transnationell styrning av utbildning och utvecklingen av nya former av expertis. Vi har identifierat detta område i våra sökprocesser, men redovisar det inte här. Däremot har vi gjort särskilda undersökningar i gråzonen – då med fokus på McKinsey-rapporterna om världsbästa utbildningssystem. Vi hade synpunkter på de argument som fördes i form av system och hur de ska förändras. Samtidigt menade vi att den betydelse som gråzonen har visar på problem att översätta forskningsresultat från fältet. Vi noterade också vi att denna typ av publikation bygger på en annan typ av expertis än den i strikt mening vetenskapliga. En viss vaksamhet är på sin plats här!

Vi försökte arbeta enligt förhärskande modeller i vårt arbete. Det visade sig att de inte sällan byggde på ett tänkande i form av orsak och verkan som inte var adekvat med den forskning om samband och mönster som vi sammanställde här. Det blev då nödvändigt att utveckla en särskilt ansats, vilket krävde en hel del tankearbete. Den typen av insatser saknas i modellen, men behövs – och kräver sina resurser – troligen inom flertalet systematiska forskningsgenomgångar inom utbildningsområdet.

En annan aspekt som är rimlig att uppmärksamma är själva översättningen från forskningens kunskapsbidrag till verksamheter inom utbildning. Vi noterade att i de genomgångna texterna betonades den politiska och professionella betydelsen av de slutsatser som dragits – och då framför allt i termer av ökad insikt än i konkreta

direktiv. Översättningsfrågan behöver emellertid hanteras på ett mer utvecklat sätt. Vi prövade att göra detta i form av undersökningar – om slutsatserna i forskningen kunde minska osäkerheten i bedömningar av vad som var möjligt eller nödvändigt att göra. I sammanhanget är det viktigt att notera skillnader mellan de-kontextualiserade slutsatser med ambitioner att dra generella slutsatser och det kunskapsunderlag som i praktiken krävs för politiskt och professionellt agerande där kontextuell giltighet är av vital betydelse. Detta gäller även för de slutsatser som förs fram i den gråzon som vi identifierade och kritiserade.

Till sist menar vi att även systematiska forskningsgenomgångar kräver att bli föremål för kritiska analyser. I vårt fall har vi kunnat notera att sådana genomgångar också bör analyseras vad gäller såväl kunskapsintresse, genomförande och konsekvenser. Vi menar att den tankestil som är för handen bör klargöras och diskuteras i relation till vilka anspråk – då också för att öka objektiviteten i den kunskap som genereras.

Särskilt viktigt vid inrättande av ett Skolforskningsinstitutet

Vår forskningsgenomgång visade på starka och svaga sidor hos systematiska forskningsgenomgångar samt på möjliga konsekvenser, exempelvis i termer av bias och i värderingar av kunskap. Detta har vi redovisat i våra slutsatser och också fört fram i våra rekommendationer.

Vad beträffar vår kärnproblematik rörande kartläggning och analys av ILSA-forskning visar vår studie inte bara på behovet av att identifiera och analysera enskilda publikationer. Vi har också kunnat identifiera ett forskningsfält och hur det vuxit fram samt vad som kännetecknar fältets intellektuella organisering – hur dess kommunikationsmönster ser ut och till vilka man adresserar sin forskning. (Noteras bör att det finns vissa skillnader mellan programmen inom ILSA).

Sådan fältkunskap om hur pass väl organiserat ett forskningsfält är kan vara av stor betydelse när man ska bedöma det underlag som forskningen kan ge. Man bör därför överväga om Skolforskningsinstitutet ska se till forskningsfält och inte bara till individuella publikationer blir föremål för systematiska forskningsgenomgångar.

Rekommendationer till Skolforskningsinstitutet

För det första är arbetet med att genomföra systematiska forskningsgenomgångar tidsmässigt och intellektuellt krävande och kräver resurser därefter.

För det andra duger det inte att komma med uttalanden om ”vad forskningen säger” eller att ge direktiv utifrån de slutsatser som dras av forskningen. Systematiska forskningsgenomgångar bör ge erbjudanden om meningsfulla samtal utifrån olika positioner som tar hänsyn till såväl vetenskaplig som politisk och professionell integritet. Då är det argumenten som ska vara i fokus.

För det tredje innebär det krav på ett tydligt och fokuserat strategiskt tänkande hos Skolforskningsinstitutet för att identifiera vilka områden och vilken forskning som ska gås igenom – och då också i vilken ordning som det ska ske. Kriterier för urval av områden bör formuleras och diskuteras med dem som berörs.

För det fjärde är det nödvändigt att ingå i eller att skapa en arena för angelägna samtal – och då gärna där kontroversiella frågor kan tas upp och där även verksamheter i gråzonen kan behandlas. Detta är ingen enkel uppgift men den måste tas itu med för att inte Skolforskningsinstitutet ska bli sittande fågel i en korseld mellan profession och akademi.

Och för det fjärde krävs kontinuerliga analyser av hur forskningsgenomgångarna fungerar och vad som behöver utvecklas hos dem. Vi hoppas att vi har visat på vissa viktiga frågor därvidlag i vår forskningsgenomgång.

Litteratur

Gough D, Oliver S, Thomas J (2013) Learning from Research: Systematic Reviews for Informing Policy Decisions: A Quick Guide. A paper for the Alliance for Useful Evidence. London: Nesta.

Hacking, Ian (1992a) “Style” for historians and philosophers. *Studies in the History and Philosophy of Science*. 23(1): pp.1-20.

Popkewitz, T. (2014). Social Epistemology, the Reason of “Reason” and the Curriculum Studies. Special Issue: Nuevas Perspectivas sobre el Curriculum Escolar. Education Policy Analysis Archives DOI: <http://dx.doi.org/10.14507/epaa.v22n22.2014>

Prøitz, T. (2015) Metoder for systematiske kunnskapsoversikter – relevant, tilgjengelig og praktisk anvendbar? Vetenskapsrådets publikasjonsserie, forthcoming spring 2015

von Wright, G. H. (1983): Practical Reason. Blackwell (1983)

von Wright, G. H. (2004). Explanation and understanding. Cornell University Press.

THE RESEARCH REVIEW PURPOSE AND PROBLEMATICS

In this chapter we depict how we are approaching the task to do a systematic review of research on international comparisons of school results. We present our purpose with this review, what questions we ask and how we reframe these questions into an argumentative framework for analysis of the field of research in context.

Comparative studies in education are of different kinds and with different connections to the Humanities and the Social Sciences. Thus, Nóvoa and Yariv-Mashal (2003) distinguish between case studies – where you compare national cases over time or over context – and variable studies – where you compare outcomes over sets of variables. The former kind has close links to the Humanities and the second to the Social Sciences, and the battles between these academic fields have also taken place in comparative studies. Steiner-Khamsi (2013), for example, adds comparison over standards to comparisons over time and context. In the comparison over standards there usually are measures of outcomes in relation to norms, such as indicators of school performances or gender equity. In such research reviews the focus is on variable research mostly dealing with comparisons over standards measured by tests on school performances, what is called International Large-Scale Assessment (ILSA).

We have defined our task as different from the above. We ask what research and with what results are carried out in this field of ILSA study? How are research results and conclusions from such comparative research carried over into educational decisions and actions?

This systematic review has as its core analysis the content of research publications and the principles that organize knowledge about school results, schools and educational change. These publications are covering different areas of interest – such as gender gaps in mathematics, efficiency in school organization or comparing school systems – and present analyses of patterns or relations between different variables. Thus, we are not dealing with specific interventions to improve schooling or the validity of specific directives. Instead, we are analyzing the production of scientific knowledge in the field of ILSA of education and inquiries in translations into actionable knowledge based on ILSA research.

Furthermore, it has to be stated that we are not doing ILSA research as such here. Instead, we are doing the review as educational researchers with the ambition to analyze educational knowledge production in research contexts and how this is translated into knowledge that is to be "*usable*" in political and professional decision-making and in the construction of expertise in education.

Background

The Swedish government decided in 2014 to implement a School Research Institute, with similar functions as the EPPI Centre in England and the Danish Clearinghouse, in order to organize and present knowledge of relevance for managing and developing education and schooling. A central task for such an institute is to produce systematic research reviews in the field of education and schooling (see Department Paper U, 2014:04) in order to further develop the scientific base for those who are working in school, e.g. with planning, teaching and evaluation. Thus, scientific quality and actionable knowledge are of vital concern for the School Research Institute. A central problem is then how to transfer knowledge from the arena of educational research to the arena of educational issues and activities.

The Swedish Ministry of Education gave the Swedish Research Council and its Committee for Educational Sciences the commission to organize and further develop systematic research reviews of importance for the foundation of the School Research Institute. As part of this process we were asked by the Committee to carry out a systematic research review of international comparisons of school results. Such reviews are conceived of as important in producing solid evidence for innovating and reforming education and in producing rational roadmaps for educational change. The introduction of specific systematic research reviews is also regarded as a criticism of current research and knowledge production for being unsystematic in their way of work.

In focus for this review are international comparisons of school results by means of International Large-Scale Assessments (here with ILSA as an acronym). The need for such a systematic research review can be understood from the increasing importance of international comparisons in current policy discourses, internationally as well as nationally (Benavot & Tanner, 2007). For instance, results from international comparisons and ranking play a very important role among political parties in recent definitions of crises in Swedish education and in discussions about strategies employed to counteract this crisis. The importance of ILSA is also evident in the statements given by the professional organizations of teacher unions and by school leader organizations and mass-media. Thus, it seems to be of vital importance to review research based on ILSA, as compiled by organizations, such as the OECD with the *Program for International Student Assessment* (PISA) or *International Association for the Evaluation of Educational Achievement* (IEA) and the *Trends in Mathematics and Science Study* (TIMSS). ILSA are a relatively new social space for thinking about education and it's planning through agencies placed institutionally outside of national systems. Yet, they are designed for national assessments and to influence planning. A systematic research review can tell us about what research is carried out and how this research might inform us about developments in the working of schools and alternative routes to take in the reforming of education and improvement of school results.

Stated otherwise, international comparisons of school results play a vital part in current educational policymaking based on measurements and numbers to react upon. Here, ILSA have become an integrated part of educational narratives in framing and reframing what education is thought to be, what the definitions of problems and tasks are, and what education is supposed to achieve (Pettersson, 2008, 2014). At the same time, there has been a proliferation of models to organize, synthesize and interpret new policy frameworks – such as international comparisons and advice from international actors based on evidence from these comparative studies. The expansion of ILSA research programs is presented in Figure 1.

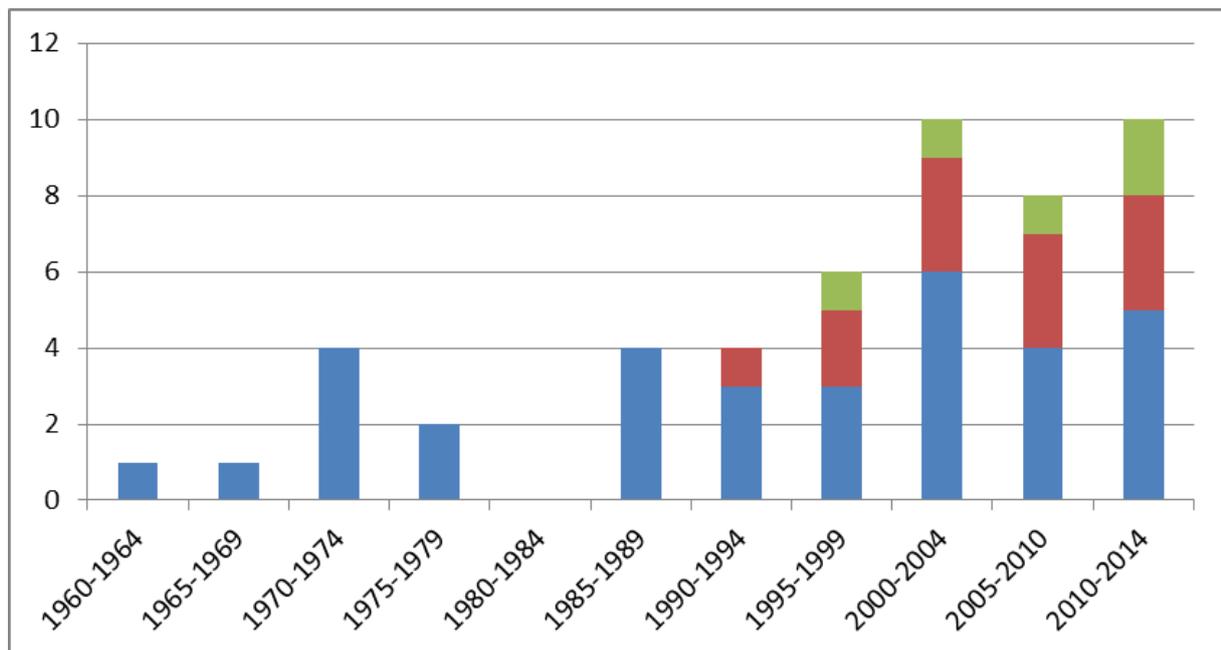


Figure 1. The expansion of ILSA studies. Blue: IEA-studies, Red: OECD-studies, Green: Regional Large-Scale Assessments. The table is based on the first year of publication of results from the various studies.

Figure 1 shows an expanding field of study. In the sixties, seventies and eighties there were only a few IEA studies. The situation changed radically in the nineties – at a time when the OECD also started to perform ILSA research. Educational achievements in different areas – mostly concerning mathematics, sciences and language competences, but also in civics – have been in focus. An indication of the increasing attention given to ILSA is the national and comparative research generated from its measurements. The number of research publications

from ILSA research itself is expanding as well. When doing our search for literature we identified more than 11 000 publications since the 1990s.

The field of study and style of reasoning

This systematic review addresses questions of the conceptualizations, organization and principles of research related to ILSA; at the same time it provides a systematical review that seeks to address the limits of current examples of policy reviews.

What does then this field of study look like? First, in the beginning of this chapter we noted the existence of different kinds of comparative studies in education and their roots in the Humanities and the Social Sciences. Our review is based on analyses of (a large number of) individual publications. Second, these publications are designed, carried out and refereed as belonging to a specific kind – as international comparisons based on large-scale assessments. Our analysis, however, considers the limits of these approaches and asks different kinds of questions in the reviews than found in the existing reviews, something we discuss later. We ask what kind of research this is, what is regarded as objective statements on education and what is regarded as progress?

In order to deal with these questions, which we think are vital in research analysis and research reviewing we turn to the concept of *style of reason* as presented by Hacking (1992a). Referring to Crombie (1988) he presents six styles of reasoning:

- a) The simple method of postulation exemplified by the Greek mathematical sciences.
- b) The development of experiment both to control postulation and to explore by observation and measurement.
- c) Hypothetical construction of analogical models.
- d) Ordering by a variety by comparison and taxonomy.
- e) Statistical analysis of regulations of populations, and the calculus of probabilities.
- f) The historical derivation of genetic development (Hacking, 1992a, p.4).

With Hacking we state this list as simple evidence for different styles of reasoning that have developed over time and are sometimes used in combination. A style of reason is important to capture in order to understand what counts as evidence and standards for objectivity and what sentences can be regarded as candidates for truth and falsehood within a style of reason.

Thus, there is circularity in the relation between style of reasoning and what is conceived of as true and false among different sentences. If we think of current reviews, they have adopted the style of reasoning that assumes the principles of the ILSAs and then proceeds to summarize and identify patterns of the ordering and classification of the given calculations. Our approach to ask about the principles of styles of reasoning around the ILSA is to understand the knowledge production and the authentications through which relevancies are established. To approach the review as styles of reasoning is a strategy to improve our understanding of the constraints and opportunities for objective knowledge produced within this style. An implication of this is that it is important to identify the style of reason at work in ILSA research.

ILSA research is to a large extent to be found in the style of (e) “statistical analysis” dealing with patterns of relations between variables and categories – based on de-contextualization and universalistic knowledge interests. Such a style has long historical roots but developed during the nineteenth and early twentieth century, marked by e.g. the emergence of probability, the theory of error and the creation of statistical objects (Hacking, 1992 b, p 141 f), serving as preconditions for producing and evaluating statistical statements, e.g. about means and deviations from that means within a population, considering for instance height or health issues. During the history of statistics we note changing ways of analyzing as well as demands of testing and intersubjectivity. ILSA research is based on the development of such statistical reasoning, for instance:

- by ways of defining and analyzing populations (comparing for instance countries or educational systems) and their characteristics
- by developing taxonomic groups (in terms of socio-economic indicators, sex, or cultures) and comparing their progress and failures in educational matters – e.g. drop out or kind of education
- by measuring means and variations in performances by means of certain tests,
- by formulating procedures or methods for the production of valid statements, such as demands on strength of association or significance tests.
- by ways of presenting and communicating research

Comparing to several other reviews, it has to be noted that research based on ILSA are not dealing explicitly with interventions, experiments and so on, often emphasized in search for evidence in systematic research reviews. What ILSA research is doing is analyzing patterns and associations between variables by using large databases, but with the unspoken assumption of the importance of these patterns and associations for educational change, as illustrated in the OECD projects of using PISA data for national educational policy reviews and implementation recommendations (see OECD - Sweden Education Policy Review, 2014).

Furthermore, while the research report formally does not claim identifying causalities and causal mechanisms (see e.g. Woessman, 2011, for emphasizing this limit) there is a practical causality in the presentation of results. Thus, a very large share of the studies that we coded are asserting a practical relevance of the research as providing highways of change of relevance for policy decision-making or professional strategies in their work.

The current research review has to consider these characteristics in the field of research, or what we consider its style of reasoning that orders how results are expressed, relationships discussed, and changes argued. Thus, we have to develop a conceptual framework for analyzing the knowledge problematics and contributions to the research field and to actionable knowledge in education.

Purpose and research questions

The purpose of this systematic research review is to conceptually present and analyze research concerning international comparisons of school results based on ILSA studies. This entails asking:

- How does one conduct a systematic research review in this field of study?
- What research is carried out and what knowledge is produced in ILSA research?
- Of what relevance is ILSA research for policy-making and professional work in the field of education?

Based on re-analyses of the activities carried out to get answers to these three questions we put forward a fourth question:

- How to examine systematic research reviews, what are significant characteristics in the review process, what are the implications for knowledge production and positioning in the field of education?

Given these questions we decided to do the research review according to what we conceived of as the principles and ways of work in this field of study. This implied that the initial layer of this review adopt meta-analytical approaches to search through vast quantities of data, to screen, code and map a huge field of research and to be highly transparent in excluding and including publications in this field. But it also meant that we had to be sensitive to the kinds of research that are carried out based on ILSA – what problematics are presented and what knowledge is produced.

The current field of research based on ILSA was shown to be huge and we identified more than 11 000 publications dealing with different problematics and research approaches. We had to find out ways of capturing this field of research in its complexity but also to do specific analyses of research dealing with specific questions. After initial searching and mapping of research publications in refereed scientific journals, we decided to start with the research-based *Programme for International Student Assessment* (PISA) organized by

the OECD. After that we turned to the IEA programs *Trends in Mathematics and Science Study* and then *the International Civic and Citizenship Education Study (ICCS)* combined with the *Civic Education Study (CIVED)* and reviewed publications based on these programs.

However, when doing so and with knowledge about current political discourses in education we realized there is a set of significant publications that intersect with the research and are mediating space between the research and policymaking, e.g. OECD national reports and the McKinsey reports that operate in a *grey-zone* between policy and research. We understand the *grey-zone* as a new space that translates research into principles for the management of decision-making and change. Given the noted impact of this *grey-zone* in forming the translation problematic that we speak of later, the inclusion of such *grey-zone literature* provides a way to think about the field and how its knowledge production, the positioning of the international reviews and expertise in educational change is recognized.

It is our overarching ambition to review the field of research on international comparisons of school performances and to analyze the contributions of a large amount of studies in the field of comparative education. Given the characteristics of this field, where neither study objects nor knowledge objects are fixed, we tried to develop a review approach where we could identify different contributions in this field – what problematics are put forwards, what facts are presented and how are these facts framed, inquired and communicated?

The research review problematics

As noted in our questions stated above, \ knowledge problems are vital in a research review. Thus, we need a tool that can be used to identify and systematize ILSA research. Stated otherwise, we need to construct a research object that can deal with the varied content in ILSA research.

Explanandum and explanans

Given the characteristics of the research in focus we choose the concepts of explanandum and explanans (in plural explananda and explanantia) derived from Georg Henrik von Wright and his work on explanation and understanding (von Wright, 1971) and practical reason (von Wright, 1983). These concepts will be developed into ingredients in a *language* dealing with research problems and contributions.

Explanandum concerns what to be explained – in the broad sense of the word – for instance, a fact or an observed behavior, while *explanans* refers to propositions aiming to explain the current fact or behavior. For instance, an explanandum could be why a person is running, and explanantia could be alternative suggestions that s/he is trying to catch a train or escaping from a dog. What is a reasonable explanans depends on inquiries carried out and conclusions based on them.

Our research review is focusing on performances in ILSA. Thus, what we have as *explananda* are such performances. *Explanantia* are then the inquiries carried out in order to understand these performances. Based on these inquiries certain inferences or conclusions are made concerning which explanans or combinations of explanantia are regarded to be reasonable in the current publication. The following concerns production of explanation in its most general form:

EXPLANATION => EXPLANANDUM x EXPLANANS => CONCLUSION

This general form could then be transformed to capture a style of reason presenting how the research problematic is framed and what inquiries are regarded as evidence.

A specific aspect in the review task concerns what is valid and useful knowledge in education. This means that knowledge acquired in one arena – educational research based on ILSA – has to be transferred to another arena – education measures and activities in specific contexts, actors and instruments. For instance, how can knowledge, based on de-contextualized comparisons with universalistic interests, be of use for acting where it is of vital importance to consider the concrete preconditions at stake?

Practical reason

A way to analyze this issue is to put it into a frame of practical reason. If we want to understand why a certain event had to – or could – happen we need to understand the interplay between contexts, actors, power relations, preconditions and expectations. We are here using practical syllogisms as a framework for analysis (cf. von Wright, 1983). A practical syllogism is a kind of syllogism, where an *action* is conceived of as necessary, given certain interests or expectations in combination with what we here call *epistemic attitude* concerning conceptions of the ability to act and constraints and opportunities at stake in the current situation. Such a practical syllogism has a general form as follows:

$$\text{ACTION} \Leftarrow \text{INTENTION} \times \text{EPISTEMIC ATTITUDE}$$

An action in a practical syllogism corresponds then to explanandum in the general form above, while explanans refer to the combination of intentions and epistemic attitudes of the actor(s) in a certain situation. Practical reason is then a *retrodition*, telling why a certain action was necessary or possible.

Given this we can combine the forms of explanation and practical reason to each other in two ways. The first is that practical reasoning is informing knowledge production in research in terms of explanandum and explanans as well as what counts as valid conclusions. The second, which is in focus here, is the other way around: how is educational research informing practical reason in education? Since we are dealing with two different arenas we talk about translations of knowledge from one arena to another, where statements are changing meaning and significance when moving from one arena to another.

Review problematics

In a way practical reasoning and research work in tandem – what is conceived of as practical problems or irritations are translated into research issues. The second is the other way around: research is informing intentions and epistemic attitudes and by that eventually governing action. This second way is in focus here. The general question is then how knowledge production by means of research is translated into intentions and epistemic attitudes: why certain actions are necessary or possible, how they can be done and what to consider when doing so. In the complex interplay in social events *practical inquiries* into research might decrease the uncertainty in concerning what can be done – and what cannot be done – in order to realize certain expectations. We explore this conceptual way of organizing the review in subsequent chapters. Based on this basic form we reformulate our questions above into three qualitatively different kinds of research problematics that we are investigating:

- *Core problematic*: What explananda are formulated in ILSA research, what explanantia are presented and what conclusions are put forwards in ILSA publications? The main point in this problematic is to capture and organize arguments put forwards in ILSA research.
- *Field problematic*: Why these core problematics? What are the reasons for formulating and framing explananda and explanantia as well as conclusions in ILSA research? Are there different positions and positionings in ILSA research? Here, the main ambition is to explain ILSA research and thus to comprehend the preconditions for the arguments presented in this field of study.
- *Translation problematic*: What is the stated relevance of ILSA research for education governance and professional work? What can be concluded from practical inquiries into research results? How is the research translated into practical reasoning and by whom?

These three problematics form the research review objects of knowledge. Thus, we are not carrying out ILSA research. Instead, we are researching what we label as a *conceptual* or *argumentative review* of ILSA research in contexts of importance for understanding this research.

ON THE DESIGN OF THE RESEARCH REVIEW

The purpose of this chapter is to present how the current systematic research review was designed and the reasons for this. The design is constructed against a background, where more principal approaches concerning research reviews are presented. Specific in our approach is that we are reviewing research that is having different study objects as well as knowledge objects. Thus, we are *not* investigating specific effects of certain educational measures or interventions or the validity of some conceptual configurations.

Based on this we present the motivation for the stages in the review and how we identify some basic problems in the making and use of research reviews in the field of education.

On systematic research reviews

There are quite a few agents in the production of systematic research reviews (SRR) since the 1970s – when Gene Glass coined the term meta-analysis. Mostly, SRR have the ambition to systematically turn primary research, such as articles in research journals, into a comprehensive picture of the state of the art of research in a specific area.

In the literatures on systematic research reviews, we find different approaches (Pröitz, 2015) and centers — e.g. the *What Works Clearinghouse* in the US focusing on meta-analyses – combining medicine and health research with public management approaches to achieve what is conceived as valid knowledge and evidence for best practice. Another example is the EPPI Centre (The Evidence for Policy and Practice Information and Coordinating Centre) in England.¹

The field of systematic reviews is based on claims that previous or traditional literature reviews are being unscientific and unsystematic – missing large fields of research and biased by researchers' own preferences, as can be read in the following quotation from Goldacre et al. (2012):

... instead of just mooching through the research literature, consciously or unconsciously picking out papers here and there that support (our) pre-existing beliefs, (we) take a scientific, systematic approach to the very process of looking for scientific evidence, ensuring that (our) evidence is as complete and representative as possible of all the research that has ever been done. (Goldacre, 2012 in Gough, Oliver, & Thomas, 2013 p.4)

On the basis of this and similar conceptions of research principles of systematic research, reviews are developed. Early, the concept of *meta-analysis* was developed by Glass & Smith (1979) with a focus on statistical studies. And later we find principles, for instance in terms of *meta-ethnography* (Noblit & Hare, 1988), where qualitative studies are interpreted and synthesized. The application fields for SRR expanded. For instance, the EPPI Centre is an example in the field of education (Oakley et al., 2005).

The competing ways of working in the field of systematic research reviews can be understood at a different layer of current scientific practices related to broad social issues and policy. They entail both demands for conceptual understandings of the fields to be reviewed and identifying what knowledge contributions are at stake. It is this layer of systematic reviews we think is important for understanding the contributions of international large-scale comparisons of school results as a field of knowledge production that implicitly or explicitly serve as roadmaps for educational change.

For example, studies in the history of science have shown how research instruments are theoretically informed (see e.g. Törnebohm, 1971) and how styles of reason are including the ways research problematics are formulated, and what is regarded as evidence is captured in this style of reason. This is not to claim that anything goes in terms of formulating research problems and outcomes. Rather it is to provide self-reflectivity

¹ <http://epi.ioe.ac.uk/>

to systematic reviews. The self-reflectivity is that research outcomes in terms of data embody a self-referential and self-authorizing style of reasoning, what Kuhn (1970) in his important analysis of science called *paradigms*. To neglect the principles that order and classify this knowledge production, shaping and fashioning the relation of the problems studied, the data that gives intelligibility to the problems, and solutions offered as change would be a serious scientific mistake. It would limit the possibilities of innovation and science in responding adequately to the social and educational issues that systemic reviews are to illuminate.

Evidence and validity of systematic research reviews

An important ambition for systematic research reviews is to provide a basis for intervention or clinical decision-making, as is shown in the field of medicine and health (for a review see Pröitz, 2015). Here, it is vital to identify what is a good knowledge basis, often termed as *best evidence*. We note here presumptions concerning hierarchies of evidence, for instance that randomized experiments give better evidence than ethnographic studies. Petticrew & Roberts (2003) present an example of such a hierarchy of evidence:

- 1) Systematic reviews and meta-analyses
- 2) Randomized controlled trials with definitive results
- 3) Randomized controlled trials with non-definitive results
- 4) Cohort studies
- 5) Case-control studies
- 6) Cross-sectional surveys
- 7) Case reports. (Petticrew & Roberts, 2003 p.527)

Petticrew & Roberts mean that such a fixed hierarchy is problematic and are instead promoting typologies for identifying strengths and weaknesses of different approaches in relation to the problem or research question to be dealt with. We share this position.

In this systematic review we are focusing on knowledge production and communication within a specific research field and how this research is translated into policy-making and professional decisions. This is a more over-embracing task which includes a variation of explananda that ILSA are dealing with. Thus, our task is to identify this research and to classify it in relation to what there is to explain as well as what inquiries are developed to do this explanation.

Furthermore, education is a somewhat problematic field of study for giving directives for at least three reasons. First, educational cases are highly contextualized. What is the case is not isolated and not easy to situate in a simple way (cf. Bernstein, 2000; Luhmann & Schorr, 2000). Second, educational phenomena are interactive "*things*," depending on how they are defined and open for looping (i.e. they are formed by how they are defined and acted upon) (cf. Hacking, 1999). And thirdly, they are often controversial, given differences in cultures or context.

Given such problematics, the task to define educational problems and solutions is complicated and so is practical reasoning in and for education (e.g. Carlgren & Lindblad, 1991; Fenstermacher & Richardson, 1993). A way to analytically deal with this is in terms of practical syllogisms, where an action is explained by a combination of intention (what to achieve) and epistemic attitude (what is conceived of as necessary to do in order to realize this intention) under current circumstances. Thus, what is important to know is the case and where it is situated in terms of interests and contexts. Retrodictive analyses of cases – how to understand why something that happened also had to happen – are ways to empirically capture practical reason at work. Policymaking or professional decision-making are examples of such cases where systematic research reviews, focused on a specific deficits or interest, are assumed to have an impact on the rationality and reasonable action in specific situations. To our understanding this is a way to capture the validity of a research review – if it is of relevance for productive decision-making.

Now ILSA as research is not dealing with specific cases or defining causal mechanisms. Instead its ambition is to identify patterns or associations between different aspects or variables in the world of education. The knowledge interest is to give accurate descriptions and generalizable conclusions concerning educational

matters. This is qualitatively different from demands on validity for action in specific cases. Given this, there is an urgent need to analyze relations between generalizable associations and valid knowledge for intervention or decision-making concerning educational affairs. For instance: What relevance is assumed in ILSA research? What translations from ILSA to practical reasoning is (to be) carried out, e.g. in terms of boundary objects, and how to estimate validity of such translations in different contextually defined cases? And given the interactive character of educational matters is ILSA research redefining educational affairs? These are questions to be debated on the basis of the current research review.

Approaches to systematic research reviews

Based on Weiss (1979), Gough (2012) put forward a distinction between configurative and aggregative approaches to systematic reviews. Where the first is conceived of as interpretative of different kinds of studies and their conceptual contributions, the second is based on aggregated collection of data from different studies. The distinction is of a heuristic kind and is constructed in order to identify choices and potential combinations in review problematics and designs based on a set of dimensions concerning aims, approach, and breadth and width, as well as structure of the review work. These approaches are summarized in Table 1:

*Table 1. Configurative and aggregative approaches to systematic research reviews.**

| Characteristics | Configurative | Aggregative |
|------------------------|---------------------------------|---------------------------------|
| Philosophy | Idealist | Realist |
| Relation to Theory | Generate | Test |
| Approach to synthesis | Configuring | Aggregating |
| Methods | Iterative Theoretical search | Exhaustive search Avoid bias |
| Quality assessment | Value study contributions | Avoid Bias |
| Product | Emergent concepts | Magnitude and precision |
| Review use | Enlightenment | Instrumental |

*The approaches are not in strict columns. Characteristics might change places in several ways.

The kind of research that we are dealing with is conceptually not fitting within the aggregative approach. However, we have the ambition to identify research publications with a similar focus in order to analyze the possibilities for capturing research results and conclusions given this focus. In that way we are designing a review corresponding to mixed methods (Sandelowski et al., 2011) approach, trying to deal with configurative as well as aggregative approaches to systematic reviews. However, our ambition is not to configure the state of the art of knowledge in relation to specific interventions or possible decisions. Instead, our intention is to identify and define the potential – or the constraints and opportunities – that ILSA research offers as a research field and in translations to educational decision-making and action.

What are then the processes at work here? We started to carry out preliminary analyzes of two systematic reviews here: one from EPPI and one from the Danish Clearinghouse. To our understanding, relevance to research questions has been of major importance when searching and scanning the field. In one of them the appraisal of research quality was put forward in the final stage.

Here we note two processes: one searching for research of relevance for the research questions, and the other a filtering process. Both are needed in order to identify and synthesize the research in focus. This meant a substantial reduction of research publications.

Case 1: The EPPI review (Gough, Oliver, & Thomas, 2013) on measuring and monitoring children’s outcomes only considered documents published in English. By means of INCA, OECD and Eurydice they

obtained 652 citations. By looking at abstracts and titles of the documents they excluded 538 of these documents and by full document screening they excluded 64 reports. The filtering process seems to be somewhat tricky, but the review ended up in 109 reports that built up the systematic map. In sum, 83 percent of the texts were filtered away.

Case 2: The higher education dropout review carried out by the Danish Clearinghouse made a more extensive search according to the sources (Sogaard et Al., 2013, p.22). They identified 6 207 unique references and ended up with 69 references – a little more than one percent of the originally found references were remaining after the filtering process.

We note here first that the research fields reviewed embrace large numbers of publications, and second that the filtering processes reduce these numbers dramatically. Taken together - transparent procedures in reviews and implications of searching and filtering research publications are vital in general as well as in our own study. As stated above our focus for reviews is conceptual, having an argumentative focus. We ask what conceptual frameworks are at work, how are different kinds of inquiries developed and what knowledge is obtained? However, we have recognized the need for analyzing the field of ILSA research and its translations into educational matters in order to understand the premises for the core research and to put it into context.

Design procedures for searching and filtering publications

Our design was based on the idea that it should be carried out according to prevailing principles and practices in order not only to present a review but also to provide experiences to reflect on the practices and principles that were used to carry through the review. These are summarized under the stages in Table 2. This reflective ambition we conceived of as consistent with the commission given to the Swedish research council by the government. Though we initially presented a review protocol, it was somewhat revised over time.

- The research review was emanating from three qualitative different but related kinds of problematics.
- The core problematic of the research review deals with knowledge production – which research, what arguments and what conclusions are presented in international comparisons of school results?
- The field problematic has the ambition to understand the production and communication of knowledge – why this research, these arguments and conclusions, and which are the different positions and interests?
- The translation problematic concerns transfer of knowledge of relevance for political and professional action. What is translated by whom and for what reasons?

Given how these problematics are handled we will have a basis for analyzing the knowledge production and contextualization of international comparisons of school results.

The core problematic

The core problematic is of primary concern for collection of data in the research review as such. Below we briefly present main ideas in this review:

First – *demands and addresses*: Of vital importance is the foundation of the School Research Institute and the policy behind it, for instance in parliamentary bills and instructions given. The idea of a School Research Institute fits well with current relevancy strategies in research policy in general and in special with current views concerning knowledge production of relevance for teachers, school-leaders and policy-makers.

Second – *scope and definitions*: Included are research publications based on international comparisons of school results in terms of ILSA. We define research publications as texts presented in peer-reviewed journals. In order to be regarded as international comparisons the publications have to be based on comparisons between at least two countries using data from ILSA and contain some kind of conclusions.

Third – *review questions within the scope and definitions presented above*: What research is carried out, what arguments are provided and how are they communicated to what addressees? Of interest is not only individual research publications but also the field of research and what relevance that is claimed.

Fourth – *the search and filtering of research*: This is done with the ambition to work in accordance with accepted principles for systematic research reviews. It is important to have a disciplined organization in the review, providing specific steps and explicit criteria for including/excluding publications and to organize the research contributions in a coherent way.

Table 2. Design for the systematic research review over stages, questions and principles for answers.

| Stage | Questions to answer – and principles in answering them |
|------------------------|--|
| Background and mission | The government and the founding of a school research institute are asking: <ul style="list-style-type: none"> • What are the demands on systematic research reviews for policy-makers and professionals? • What are the uses of international large-scale assessments for these actors? • What are the opportunities and limits for systematic research reviews? |
| Review Questions | <ul style="list-style-type: none"> • Identifying: What research on international comparisons of school results? • Explaining: Which arguments concerning differences and similarities of school results? • Concluding: What insights are presented and directives are proposed for what reasons? • Reflecting: Implications of systematic reviews for research and knowledge production? |
| Scope | Definitions and criteria for including research: <ul style="list-style-type: none"> • Studies using international large-scale assessment data • Two or more countries are compared • Explanations of results are presented and/or directives to improve results are given |
| Search | Search is developed by means of Discovery complemented by specific searches in Web of Science, Google Scholar and ERIC |
| Screen | Develop criteria for which studies to include. Filtering strategies: <ul style="list-style-type: none"> • Classification of studies in accordance with criteria • Selecting studies in accordance with inclusion/exclusion criteria |
| Code | Presentation of definitions and criteria. Coding of relevant aspects for research: <ul style="list-style-type: none"> • Inquiries • Research questions • Types of account • Translation research results to issues of practical relevance • Assumptions of relevance |
| Map | Categorization of explananda and explanantia. Construction of interesting subgroup for further analysis and synthesis. |
| Appraise | Assessment of quality. Assessment of relevance. |
| Synthesize | Answers to research questions over problematics: <ul style="list-style-type: none"> • Core problematic and Review questions • Field Problematic • Translation problematic |
| Communicate | Present research in different stages: <ul style="list-style-type: none"> • Commission groups meetings • National conference • International communication |

We focus on the OECD and the IEA studies and research based on the databases provided by these organizations. The identification of research references will be done by using the over-embracing search tool Discovery as presented below. The screening process is preceded by filtering processes according to specific criteria presented and operations carried out in relation to this. This is presented for each program.

Our design is based on a combination of different approaches. Partly, our work is a review of the research field where we use a design in accordance with an aggregative approach with exhaustive search and transparent filtration, and partly it has characteristics similar to a configurative approach with a focus on concepts and the development of arguments.

During the coding and mapping stages we developed what we named an argumentative approach in inquiries dealing with the interplay between *explanandum* – facts or results – and *explanans* – how these results or facts are understood or explained and what conclusions are made in these inquiries (cf. von Wright, 1971).

Coding of each selected research publication is carried out using a specific sheet developed for this purpose. The coding is based on identifying inquiries and research questions in combination with types of research accounts, research relevance issues and conceptions of translations between research results and educational practice. Coding content and summaries was developed for each of the research programs and their selections of publications. After that a mapping took place based on the coding but organized over the explananda in the publications. The mapping turns the publications into a language of explanandum, explanans and conclusions. During the synthesizing stage our manuscript was reviewed by and discussed with experts who have been engaged in the McKinsey and the OECD research programs.

The field problematic

In order to analyze the field problematics we carry out three activities for each of the analyzed ILSA programs. First, we coded the filtered publications in relation to how the authors presented their work – the inquiries and results in relation to the addressees. Second, we analyzed the publication patterns by means of a bibliometrical tool – in which journals do we find the publications and who is citing whom. These bibliometrics are presented in a specific chapter. And third, we investigated the history of the international research programs over time, actors and organizations during 1960 – 2014 and related this to the trajectories of comparative studies in education.

The translation problematic

The translation problematic is vital in systematic research reviews aiming to have an impact on political and professional decision-making. Here, we turn to the reviewed publications and code their presentation of the relevance of their research for addressees outside academia. We also analyze a selection of texts in the grey-zone between research and policymaking. These texts are selected since they are recognized as highly relevant in public discourses. Here, we analyze the arguments in order to capture how the translation problem is handled.

Search tool: Discovery systems in systematic reviews

Difference between “discovery” and “search”

Normally, traditional library information systems like library catalogs and databases are referred to as “*search systems*.” These systems offer structured information that has been tailored to the data they hold. The records are normally homogenous, meaning that they are cataloged in the same way, having the same data structure and normally relate to one single topic. These more traditional systems typically expect users to have medium to high searching literacy (Sadeh, 2013). The rise of the next generation’s catalogs was introduced in 2007 with a significant change in the way libraries engage with users. The new systems came to be called discovery-and-delivery systems or just, discovery systems. This system was more user-friendly in that it included institutional repositories, course materials and even library web pages in one single search. The development was rapid, and

already in 2009 it became possible to serve users an entire library collection from one single point of access. The discovery system solution was rapidly picked up by a lot of agencies, and during a short period different solutions were presented - such as Ex Libris Primo, Serial Solutions Summon, OCLC WorldCat Local and the discovery system used for this report, EBSCO Discovery Service (Sadeh, 2013).

One of the major concerns of users in the traditional search era was fragmentation of the search scope. For satisfying researchers' informational needs they had to search in various scholarly systems – such as library catalogs, digital repositories and third-party databases. As a result of these difficulties meta-search systems that addressed the search capacity in multiple resources were introduced. One of the major goals of discovery systems was to develop central indexes of global scholarly materials. Through this the discovery systems were able to serve users from one single point and gave access to far more resources than before. It also frees users from having to select the most appropriate information resources for searching, especially if the information needed is interdisciplinary (Sadeh, 2013).

Search with discovery systems

Discovery systems were designed after the Google machinery already had set the standard. Google as a search engine accommodates a trial-and-error approach, meaning that the interaction with the search engine is so simple that if the typical user is not satisfied with the result, he/she just rephrases the query and tries again. Discovery system developers aimed to provide the same experience (Sadeh, 2013). Because of this user-friendly environment, it was easy for us to use different “*search terms*” in finding what was most appropriate for our systemic review. Several factors besides the user-friendly environment contributed to our choosing of EBSCO Discovery Service. The first and most obvious is the fact that we have access to this discovery system. Another fact is that the content indexed is formatted much better than those indexed by other search engines, such as Google Scholar. Such is the case especially when it comes to content that consists of bibliographic records and textual information like abstracts and full texts.

There are some restrictions however in the EBSCO Discovery Service we used. The restrictions and difficulties hardest to handle in the process are that the register of subjects seems to be less developed, the restriction to peer-reviewed articles seems to embrace some difficulties and the list presented involves duplicates because if an article is available in several databases it might appear several times. These restrictions had to be handled in some way.

Our way to deal with these difficulties was as follows: i) the way to handle the lesser developed register of subjects came to be that we included all the subjects listed in the basic search and manually singled out the articles misplaced; ii) the problems with singling out the peer-reviewed articles was a more time-consuming process. We had to manually read through all the abstracts and find a system of articles being classified as peer-reviewed, and articles to be classified as newsletters, editorial or *grey-zone literature*. In our listings of these articles we merged these groups into one single classification of newsletters; and finally, iii) the duplicates had to be removed manually from the list. Another peculiarity with EBSCO Discovery Service is that when doing a search, one amount of articles is given but when going through the available list; some of the duplicates are removed automatically when reaching the end of the list. In that respect the amount is not accurate in relation to the actual amount of published articles.

However, the databases included or excluded in EBSCO Discovery Service is the most obvious restriction. In the appendix we put in a table for presenting the databases included in the discovery system. In addition, we used Google Scholar, ERIC, Swepub, Libris, DiVA and the ISI Web of Science to complement the Discovery Service. Here we will be careful to note the implications of searching and filtering in the classification of studies.

New kids on the block

When investigating international comparisons of school results the movement of educational ideas, policies and practices from one place to another has to be addressed. In relation to these questions, considerable research has been done in the field of comparative education, especially discussing these issues within the concept of

transfer. In a situation of “globalization” these questions have become more complex. Issues like what is national and what is global in education, what is inside and what is outside, and who are the actors and under what circumstances do they act have become really important questions to find answers to. In this situation global agencies have come to play a more important role. These organizations are widely spread and are disparate in aims, methods and concerns. Some are directed jointly by nations; others are directed independently. Some are open with where the resources are allocated; others are mute on these issues. In this review we primarily deal with three organizations that are influential in contemporary scientific discussions about transfer and governance – IEA, OECD and McKinsey. There are also other organizations of importance (e.g. UNESCO, World Bank), but to limit the review we are using these three as examples of the *new kids on the block* in the global educational neighborhood.

There are differences between IEA, OECD and McKinsey, but there is also a variety of similarities. The most obvious similarity is that all three – albeit in different ways – produce texts for soft governance and policy directives on education. IEA and OECD produce ILSA, which McKinsey does not, but McKinsey is very keen on using the results in different ILSA for addressing various issues in education. All three organizations produce results, discussions and guidelines on how to address educational questions. In this review we categorize these texts as *grey-zone literature*. By that we mean that even though the texts use research-like methodologies and theories, they lack in some other aspect on what research might be defined as. The most evident are i) there is sometimes trouble in recognizing where the financial resources are allocated, ii) there is also sometimes trouble in recognizing the authors of the texts, iii) often it is hard to recognize if the results have been under peer review or other kind of quality assurance outside the own organization, and finally iv) the texts lack in critical perspectives when discussing results and the questions asked. Despite these shortcomings we cannot ignore these products when we analyze what is normally discussed within the field of school results. Our statement is that these products play an important role when discussing school results on the international and national arena, but we cannot include them as scientific research publications, and because of that we use the description of *grey-zone literature* when discussing them.

With this as a point of departure we conclude in that there are some differences concerning levels of research. The review implicitly discusses the *grey-zone literature* when we describe the field, except when it comes to McKinsey. McKinsey is recognized for being a relatively new actor of importance. The organization does not itself produce ILSA but is most active in discussing the matters fore-fronted in different ILSA. In that respect we find it important to discuss McKinsey as a newer actor in the field of transfer.

The main focus in this review is the research produced in peer-reviewed scientific journals. In this we can see a practice on how *grey-zone literature* meets research and in this we can reach some overall conclusions on what kind of research the ILSA leads to and how it is translated in e.g. the political and professional arenas.

From the beginning, it is important to already state that the *grey-zone literature* in no way is “bad” or “wrong” in how it interprets and integrates research with the development of intervention models: it however operates within a space that basically does not follow the process of quality assurance outside the organization itself. Another important statement that also has to be given at the beginning is that the peer-reviewed research we study in our review is connecting to the *grey-zone literature* for addressing questions, gathering statistics and results, and doing the research manageably.

One evident example of how this *grey-zone literature* frames questions to be asked within the scientific community can be visualized with the help of one publication source within OECD, called *PISA in Focus*. From the first PISA study, the organization has published 44 shorter notes in the form of newsletters addressing what the OECD narratively creates as important questions to be asked in relation to the results and statistics emanating from the PISA studies. The caption is normally formulated as a question, and if these questions are studied one can easily follow the questions asked within different PISA studies, depth reports and research papers. The form of these newsletters is normally well equipped with pictures, tables and statistics. But they also contain descriptions on where more research and in-depth studies are needed. Some examples of these captions are *how do some students overcome their socio-economic background? What do students think about school? Do parents’ occupation have an impact on student performance?*, or, *Are boys and girls ready for the digital age?*. The content of these newsletters is normally focused on experience gaps between different groups of students, normally differentiated by socio-economic background, gender or attitudes toward school or

schooling. As such, a self-referential and self-authorizing system becomes visible. OECD, with the help of PISA, is creating specific focuses within their studies, reproduced and reframed once again with the help of the reports, in-depth studies, but also - as the exemplification that *PISA in Focus* shows – within different newsletters for promoting the studies that promote the use of PISA data within policymaking while also directly or indirectly catering to the scientific community.

However, the main objective of this review is to analyze ILSA's impact on the international scientific community. This impact can be analyzed by different means. In our review we focus on the analysis of the results of research by means of the analysis of articles published in international peer-reviewed journals. The methodological approach is based on bibliometrics, which is a quantitative analysis of publications to describe patterns within the corpus of the field. The limitations within these kinds of studies are well known and because of this we also perform a qualitative analysis of the articles in use. As such we can present measures of the impact of ILSA in the scientific community of peer-reviewed journals. Before we move to a more hands-on analysis, we will describe how the foundation for ILSA was created and how the first ILSA was established and disseminated.

A HISTORY OF COMPARATIVE EDUCATION AND PROGRAMS FOR INTERNATIONAL LARGE-SCALE ASSESSMENTS

The transfer and circulation of ideas in relation to the onset of mass schooling in the nineteenth and twentieth century led to a general curiosity about other countries' educational processes. A range of activities, such as international missions, organization of exhibitions and production of international encyclopedias, led to the popularization of comparisons (Nóvoa & Yariv-Mashal, 2003). In parallel, national measurement of education became a defining element in the governing of education, with close connections to intelligence testing and factorial analyses (Lawn, 2013). At first for internal usage, e.g. for budgetary reasons, eventually it widened to include comparisons within nations and later on between nations (Landahl & Lundahl, 2013).

This chapter is about the two phenomena, educational comparisons and data usage, and how they inspired each other. After outlining a short history we go on to discuss how ILSA was created. Relating to both phenomena, we argue that particular configurations in scientific development have enabled researchers to address and develop ILSA within a comparative discourse; thereby creating specific types of educational narratives (cf. Pettersson, 2014) and styles of reasoning (Hacking, 1992a).

A short history of comparative education studies

Comparing education systems and organizations

Comparing education is as old as the hills, so is also to count people and holdings in administrative purposes for military service and tax surveys. For western societies this can be dated back to at least the days of William the Conqueror and the administration of Domesday Book of 1086.² The survey's main purpose was to determine what taxes needed to be paid based on how much land and livestock each landholder had. Another more modern example on how to use an archaic version of statistics is when the US government in 1780 initiated census data. Throughout history, travelers have visited other civilizations and returned with facts and impressions about foreign cultures and education. These traditions developed at least through three different lines: i) comparisons between the contemporary and the history, ii) comparisons between countries and iii) comparisons for implementing "best practices." An example of the former is when Jean-Jacques Rousseau published *Émile* in 1762. In the aftermath conservative educators were dismayed because they saw a threat to the traditional pattern of education. Accordingly, they attempted to discredit Rousseau by pointing out limitations in the light of pedagogical values and content, stretching back to ancient history in an attempt to determine whether the past or present was best (Brickman, 1960).

Another development is comparisons between countries. An early example is when Friedrich August Hecht in 1795 compared schools in England and some German states. The major content is plain descriptive texts, but some comparisons were also made. Hecht initiated a German discussion about the differences between *Auslandspädagogik* and *Vergleichende Erziehungswissenschaft*. Even though the former is normally given a lower status, collecting educational data from abroad cannot be overlooked even though analytical descriptions may be lacking (Brickman, 1960). However, a clearly defined line began to develop considering differences between the study of education in other countries and studies that compared different national educational systems.

² The Domesday Book is a survey written in Medieval Latin. Today the manuscript is held at the National Archives at Kew, London. In 2011 the Open Domesday site (<http://www.domesdaymap.co.uk/>) made the manuscript available online.

During the nineteenth century the desire for educational reforms led to a quest for information and data concerning education in other countries. A multitude of literature developed with an interest in describing different educational systems. On the one hand this literature was written by travelers with an interest in education and on the other by government officials collecting data for governmental reports. At the same time, writings about foreign practices were published with the implicit intention of borrowing ideas for domestic use. However, we have to conclude that the belief in the truth-telling capacity of numbers, in order to establish values about social and personal life, has not been the case before the 19th century. Prior, truth was established through the manners and rhetorical qualities combined with social status of the speaker (Poovey, 1998). In times of social turmoil in the aftermath of the French and American revolutions moral, progressive and scientific campaigns appeared to ameliorate bad social conditions. In particular, this is envisioned within science, which during this period took an empirical turn in order to understand general social processes. By turning into a more empirical approach science could distance itself from the moral and progressive value laden, social activist roots and by that give itself an appearance of neutrality. One early example of this was when the Chicago School of Sociology inspected fragments of industrial society to understand the processes in neighborhood transition they refrained from making claims about American life at large. In doing so they were able to establish a feeling of neutrality and a more modern vision of scientificity.

Numbers and comparisons – style of reasoning

As stated above, the empirical turn led to the emergence of a new scientific branch which we nowadays know under the description of statistics. The usefulness of statistics led to that "*facts*" about social life became part of a societal change that traversed different sectors like economy, statecraft and culture during the 18th and 19th centuries. Two early examples (pre-empirical turn) of the emergence of statistics (more correctly described as collecting numerical data) is how the British government collected a considerable amount of numerical information during the three first quarters of the 18th century and the previously mentioned US government census data ordained in 1780. These data were however not collected in the context of coherent theory about statecraft (Poovey, 1998). An explanation to this might be that the numbers collected were devaluated through the priority given to the Newtonian universals and the invisible laws of nature. Consequently there was no interest in combining the individual data into cohorts in an attempt to say something about society and the societal priorities. One striking example is Sweden where numbers were used as an official part of governing through registering the reading ability of the population, but, the register was individual and without any reasoning in ordering the population that is so much more apparent in the 19th century.

The term "*statistics*", as a numerical expression trying to capture human activity, gradually emerged from the 18th century German descriptions of "*cameral statistics*" as a scientific description of varied aspects of the state. Later and through successive (re)visioning of the word "*statistics*" it came to separate into two different sectors or conceptualizations: one describing the political management of people and one describing the scientific management of phenomena and the autonomy of statistics as a separate field of knowledge. Desrosier (1991) describes these changes in the terms that statistics in the 18th century was primarily a literary term, during the 19th century it transformed into being a numerical description of the state, and again, transformed in the 20th century into interrelated or tied up to, the mathematical techniques for numerical analysis of data. As such, statistics became the node in describing or explaining the world. Today, in the 21st century it might be the place to wonder what the contemporary conceptualization of statistics might be. Is it our way to understand life? Is our understanding of knowledge even possible without statistics? Is statistics our foundation in a "*modern*" style of reasoning?

Poovey (1998) makes an argument that faith in numbers as a modern "*fact*" is found, at least in part, with the emergence of commerce. As an example, double accounting is uplifted as an aspect of how to install trust independent of the speaker and his/her social position. And instead numbers are used as "*facts*" independent of these other influences. The double accounting is what we today think of in terms of the checkbook's ledger where money received and paid out is entered. The activity of double accounting came to diffuse and mutate in uneven ways into the science of political economy and moral philosophy from the 1790s and onward. The early British theorists of wealth and society developed a mode of analysis that basically had no need for numerical

data. Later, for instance in the economic theory of Adam Smith, numbers appeared as a strategy to actualize the philosophized fictions of markets as performativity standards instead of just plain descriptions. As such, questions can be asked like: Are notions about markets possible without numbers? Are performativity standards possible without markets? To what extent have numbers created our fictional view of the world as different markets?

Desrosières (1991) argues that statistical objectifications contribute to both institutional and cognitive consolidation of social science with the help of certain pre-constructed equivalences for society while also constructing the equivalence itself. The international comparative measures we are concerned with are one evident example of this process. Statistical comparisons create recognition of differences between nomenclatures as a problem that must be eliminated. In doing so a grid can be constructed that appears to be valid and unresponsive of national contexts or time. As a result of this comparative attitude, information about contemporary taxonomies is preserved instead of dissolved. This view, discussed by Desrosières, also marks a clear rupture with the more classical ways of social science where numbers were used to describe things that exist independently of the conventions establishing them. With the entrance of a new scientific and political way to use numbers in terms of statistical descriptions, it becomes evident that the coding itself creates equivalence. The act of coding came to construct equivalence classes between diverse objects, and the class more than the individual objects came to be judged and described. In that way objects, through the process of constructing equivalence classes, were made comparable. The effect of this development was that the individual was lost in favor of overall descriptions, and numbers were then used for describing overall emphasis more than an emphasis on the individual. However, the coding to find equivalence also provides a tight link between political and cognitive dimensions. One of these tighter links can be observed in the 19th century when equity and equivalence appear as prescriptive and descriptive aspects of disciplines. Initially, comparability and equivalence between objects were less a question of knowledge and more of justice, regulated in e.g. laws governing market exchange, and as such more focused on quality aspects. Quantification through numbers (in terms of archaic statistics) had emerged in the 18th century in the German states for separating the act of political management of people from the scientific management of things. In the evolution of this separation, transcendence is taking form concerning the contingency of particular cases and circumstances to things that hold together and displays qualities of generality and permanence. The development raises a lot of question and one of them is if these objects really are equivalent, but maybe a more appropriate standing is who decides to treat these objects as equivalent, and why? As such numbers can be seen as a technology of distance used as a claim of objectivity instantiated by moral and political discourses (Porter, 1995). All kinds of quantified knowledge are in that respect artificial through creating uniformity among different qualities of things, uniformity that gives social authority to the interrelation of science and policy.

The use of numbers in social science was not a central aspect to the late 19th century scholars, also including economics, because the researchers were more speculative than empirical. The insertion of statistics into social theory had an effect in reducing what up until then was seen to be uncertain. Statistics in this historical context, to borrow from Hacking (1990), tamed chance. Statistics in that way came to give stability to things in flux and inscribed an apparent consensus that made things of the world seem amenable to control. Therefore, with the help of statistics, notions toward decision-making, human interest and problem solving are ways to order and regularize the processes in the world where the future has no guarantee, only conditionality.

Numbers became visualized as social facts whose objectivity and impartiality were important in the making of citizens in the 19th century. As such, numbers can be thought as a social technology that seems to instantiate a consensus and harmony in a world appearing otherwise: uncertain, ambiguous and contentious. The uniformity given by numbers brings order in social life by regulating relations among social and psychological components (Rose, 1999). By correlating the statistical magnitudes of these characteristics of populations to achievement levels of children a more equal and democratic society is thought to be achieved. In that way numbers perform as technologies to chart and compare populations that attest to the “truth” of the qualities assessed. Mapping of boundaries and internal characteristics of spaces to manage was therefore a strategy to make judgments outside the apparent subjective. However, while the things of numbers “act” as real, they embody implicit choices about “*what to measure, how to measure it, how often to measure it and how to present and interpret the results*” (Rose, 1999, p.199).

The apparently quantitative precision and specific delineation of social and personal life lent authority to the emerging new regimes of government connected to modernization. Modern nations organized around the idea that the citizens would participate in the “*life of democracy*,” even though the very regimes associated with democracy entail uncertainty. For example, the futures of modern liberal governments are tied to the reason and actions of individuals whose participation is to secure the common good. A theory of agency and action becomes in that respect central to participation, change and the future functioning of government itself. In that sense, the future becomes a normative part of the present and traditions questioned in relation to giving directions to progress. In this context of uncertainty new forms of government might seem oxymoronic, but that might not be exactly historically correct. Both the American and the French revolutions required systems of social administration that enabled production of particular kinds of human characteristics that would embody the collective norms and values assigned as civic virtues enacted through the individual citizen’s self-responsibility, self-motivation and participation.

A side-effect of the modern state and the use of numbers was the creation of the “*average citizen*” of a nation. By creating this “*average citizen*” a study could distance itself from the social problems related to different individuals and instead create a credo for how e.g. a particular product appeared in the public eye. Polling became in that sense a tool for scoping what common persons think. This was made through tracking small samples to stand for the whole public opinion. These pollings were not directed to reforms as earlier, instead they were directed against specific social problems or specific affections of different objects. The claim was in that respect to not just stage research on specific objects, instead they were directed against making claims on total populations. As such, one of the first and most well-known researchers, Gallup,³ saw himself as a representative of the “*people’s voice*,” measuring public opinions. This was done with an idea of looking at the masses as a giant laboratory of the mind, and related to this it was possible to build a neutral vocabulary for comparisons. Igo (2007) argues that this scientific service of staging polls created a new way of comparisons that not just created a modern culture; it also created a way for people to understand the very culture. In the early 1950s new techniques of polling, sampling and quantifying were bound up with the specific idea of citizenship. The surveys demarcated norms of inclusion, exclusion and affinity for citizens to see themselves as part of a new collective, a collective constituted by and reflected in the data compiled by anonymous others (Igo, 2007).

In sum, it can be stated that for understanding these qualities of governing we first need to consider numbers as defining a problematized space where subjects and objects are stabilized. Numbers seem technical, objective and calculable and embodying the idea of giving all equal chances and representation. Numbers standardized the subject of measurement and the act of exchange so that they were no longer seen as dependent on the personalities or the statuses of those who performed the measurement. The faith in numbers in social affairs today is so markedly part of common sense that it is possible to talk about “*transparency*” in governmental social affairs and even personal relations can be discussed through statistical charts and graphs. Making government “*transparent*” have become an act of democratic modes of acting for ensuring that everybody “*knows*” how decisions are made. In that sense, numbers have become part of a narrative about guaranteeing democracy. Consequently, the word “*transparent*” conjures an image of distance, impartiality and a mechanical objectivity that ensures “*the people*” to know and understand governing and gives them the opportunity to act wisely as citizens.

Methodologies for comparisons: Two trajectories

The notion of comparison is in one respect not new and in another respect something that has particular historical qualities that can be identified in the Enlightenment. Giving humanity its own history and also the problem of progress that was tied to an idea of cosmopolitan reasons and science was also a new mode of thought that could look at distant places and people to compare them in a manner to differentiate. The quarrel

³ George Gallup (1901-1984) was an American pioneer of survey sampling techniques and inventor of the Gallup poll.

of the Ancients and the Moderns embodied this notion of comparison, asking if the present superseded the past or merely building on what others had done. And with this comparative mode of thought was also comparing people along a continuum of value to differentiate advanced and less advanced civilizations of what Europe was the model. It is, we suggested above, that this notion of comparison as something to measure through statistical techniques applied to groups rather than individuals appears in the late 18th and early 19th century in education.

The French scientist Marc-Antoine Jullien was one of the first to construct a methodology for comparisons that: i) separated the empirical field of observation into its constituent parts, ii) devised techniques of inquiry, and iii) used formal models of analysis in an attempt to explain how schools functioned (Gautherin, 1993). In 1817, Jullien tried to compare educational establishments throughout Europe by setting up a *Special Commission on Education and an Educational Institute* publishing *Educational Newsletters* (Brickman, 2010). As part of this project researchers were sent to different countries to investigate their educational systems with the aim of identifying the best schools. These studies, which were empirical in approach and implicitly comparative, were intended to identify aspects of education that could easily be imported into the French system of education (Gautherin, 1993). Jullien was convinced that collecting facts and observations arranged in analytical tables would advance the science of education in the same way as the natural sciences. He thought that he would be able to find laws that governed the characteristics of education. With his comparative approach Jullien imagined that he had found an instrument that could be used to analyze observed facts and facilitate investigations. He also needed a tool that would create a comparative table of the main educational establishments in Europe, show the organization of education and the methods used to instruct students, and that indicated improvement and success rates (Gautherin, 1993). This tool was provided by a standardized questionnaire of 266 questions published in the scientific journal *Journal Dn* (2010). Jullien hoped to collect a variety of observations that could be compared by means of a common measure used for correlation (Gautherin, 1993). Jullien's vision of an educational science was thus much broader in scope than any other. However, he was also looking for practical benefits that would make it possible to judge which improvements could be transported from one country to another in order to improve a country's internal methods. From Jullien's point of view education was seen as an independent aspect of social reality that could be analyzed separately from the socio-historical context (Beech, 2006). In this, educational transfer was seen as a desirable process that could give birth to the idea of borrowing what was seen as good and useful in order to develop a better education (Fraser, 1964). Furthermore, he thought that general educational principles could be deduced and applied to improve education in most contexts, and that when universal ideas of education were established from a series of comparative tables it would be possible to use this model to make improvements (Fraser, 1964).

Apart of Jullien's ideas about a search for general principles and the creation of international agencies, the nineteenth century was dominated by men appointed by their governments to develop internal education (Holmes, 1981). Jullien's work went more or less unnoticed during this period, and it is even appropriate to talk about a rediscovery in the late nineteenth century. As such, Jullien's methodology was not especially important for comparisons during the nineteenth century. It was only when comparative education as a scientific branch grew in importance in the late nineteenth and early twentieth century that Jullien's work was discussed and appreciated. Another practice evolved too – the increased use of national data in education systems (Lawn, 2013).

Exhibitions and scientific congresses of the nineteenth century fostered a specific discourse in relation to numerical data. Numerical data was used for comparisons between different educational systems. The first *International Statistical Congress* was held in Brussels in 1853 and already from the beginning education was seen as an interesting area for statistical comparisons. These congresses were held on a regular basis until World War I and as such became important nodes for educational data. Many of the studies conducted were national in nature although some did focus on international comparisons (Smyth, 2008). The growing authority of data in governing education systems during the nineteenth century was mostly influenced by the use of data in the US. *The Department of Education* was created in 1867 and later reorganized as the *Bureau of Education*. The agency's main task was to collect and disseminate educational statistics although the data was also seen as a catalyst in the improvement of education (Lawn, 2013). During this period the most famous reports were

those written by the Secretary of the Massachusetts Board of Education, Horace Mann. He had in 1838 started to publish reports about American schools, and in 1844 he published a report (Mann, 1844) in which the Massachusetts school system and schools in Great Britain, France, Germany and Holland were compared. Mann's report served as a model for such comparisons and had a major effect on other reports of different educational systems.

When an exhibition was held in Paris in 1878, the US was able to produce a variety of reports using educational data, which in turn influenced the future of comparisons (Lawn, 2013). Especially important was the numerical data visualized as graphs and diagrams. Graphs and diagrams could store large amounts of data and make explicit claims on data relations that were easy for people to grasp. These images were able to “*speak for themselves*,” as Lawn (2013) puts it. The statistical data that was displayed in graphs and diagrams was also standardized, thus creating a new language and interpretation of society (Lawn, 2013). The practice of using statistics and presenting results in graphs and diagrams spread rapidly and became part of the educational language. In Porter's (1995) words, using the “*language of quantity*” has the advantage of being “*the technology of distance*” (a.a., p. ix) and a decision based on numbers has at least the appearance of being fair and impersonal.

Spaces for comparison and data: From comparing to collaboration and learning

The interest in collecting educational data increased and led to systematic comparisons becoming relatively common. Examples can be found in governmental organizations such as the U.S. *Bureau of Education* (1867), the *Musée Pédagogique* (1879), the *Office of Special Inquiries and Reports* in London (1895), and the *Zentralinstitut für Erziehung und Unterricht* in Berlin (1915) — all of which collected data in order to develop education (Brickman, 1966).

In 1900, the comparative educationalist Michael Sadler delivered a speech called *How Far Can We Learn Anything of Practical Value from the Study of Foreign Systems of Education?* This more or less signaled a new era of comparing educational systems – a period in which researchers like Isaac Leon Kandel, Friedrich Schneider, Nicholas Hans and others laid the foundations and built structures for the scientific field of comparative education. In the first decades of the twentieth century monographs, yearbooks, statistical compilations, conferences, instructions, and discussions about practical applications of comparative education started to multiply. Even though statistical data was evident during this period the foundation of comparative education was mainly rooted within humanities, but eventually the scientific community started to make arguments for stronger educational planning and the development of means for reliable predictions (Noah & Eckstein, 1969). As a result, independent organizations were developed in order to compare the education provided by different nations using the available data. These included the *Institute of International Education* in New York (1919), the *International Institute of Teachers College*, Columbia University (1923), the *Bureau of International Education* (IBE) in Geneva (1925) and the *Institut International de Coopération Intellectuelle* in Paris (1925), (Brickman, 1966). Connected to these institutes were a variety of scientific journals, such as the *Educational Yearbook* from Teachers College, Columbia University, edited by Isaac Leon Kandel between the years 1925 and 1944 and having a major impact in the areas of comparative and international education; the *Year Book of Education* (1932-1940) published by the University of London; the *Annuaire International de l'Éducation et de l'Enseignement* issued by the International Bureau of Education between the years 1933 and 1939 (Brickman, 1966). Of particular importance in this context was Professor Kandel's seminal book *Comparative Education* (Kandel, 1933), which was translated and disseminated (Brickman, 1966).

After World War Two the field of comparative education developed in somewhat different trajectories – more into collaborating and learning from each other. Some of the older institutions for comparisons were revitalized, such as the Institute of International Education and the *International Bureau of Education* (IBE).

New ones also appeared such as UNESCO (1946), the *Institut für Vergleichende Erziehungswissenschaft* in Salzburg (1946-1953), the *Pädagogische Arbeitsstelle* in Wiesbaden, later in Bonn (1947), the *Hochschule (later Deutsches Institut) für Internationale Pädagogische Forschung* in Frankfurt-am-Main (1949), the *UNESCO Institut für Pädagogik* in Hamburg (1951), the *Center of Comparative Education* at the University of Ottawa (1954) and the *Research Institute of Comparative Education and Culture* at the University of Kyushu, Japan (1954) (Brickman, 1966). A number of conferences

were held with a view to strengthening the field and putting more emphasis on collaboration and learning than comparison.

In short it can be said that comparison is a historically well-known fact. Historically, there have been attempts to develop methodologies and collaborative networks. After 1900 the scientific field of comparative education strengthened and was able to address questions concerning methodology issues, institutions, organizations in journals and seminars discussing how and what to compare. The collection of educational data developed at the same time. At first this was mostly descriptive data relating to different educational systems for governance reasons although there was later a shift towards a more knowledge output-based discourse (Landahl & Lundahl, 2013). These two parallel processes of comparison and the collection of data for educational purposes (Lawn, 2013) strengthened each other and created a specific discourse, narrative or style of reasoning on education that made it possible to stage ILSA of student achievement.

A history of large-scale assessments

Performance in schools has during the last decades been increasingly judged on the basis of effective student learning outcomes. Countries inspired by the importance of comparisons and data developed a range of tools and techniques for evaluation and assessment as part of their efforts to improve student learning outcomes. This happened because education has been characterized as a central requirement for national economic development and political democratization. A way to deal with this requirement is through benchmarking, identified as the “[...] *basis for improvement* [...] *It is only through such benchmarking that countries can understand relative strengths and weaknesses of their education systems and identify best practices and ways forwards*” (OECD, 2006, p.18). Another slogan is encapsulated in the World Bank statements saying, “*Examine, assess and compare*” (World Bank, 2005). Statements like these exemplify the international consensus where assessments and comparisons are highlighted as a necessity. However, assessments are often linked to efforts to reform educational systems and are themselves stimuli for further reforms (Baker & LeTendre, 2005). One aspect of this development is the increasing participation in international, regional and national knowledge assessments.

Since the end of the nineteenth century the production of numbered data starts to be used in at least the western countries for bringing new visions of the social and economic world. The new construction of epistemic references for defining “*reality*” with the help of data is linked to the creation and management of the development of the self-defined “*democratic*” state. Numerical data also provided more than an “*objective way*” of seeing reality, it “*instituted*” the reality by creating a “*common cognitive space*” that could be both observed and described through data (Lussi, Borer, & Lawn, 2013). This “*common cognitive space*” has also been framed within the narratives of different international, regional and national assessments.

International Assessments

After the Second World War, data were gradually considered more as the most objective way to understand “*reality*” (Lussi, Borer, & Lawn, 2013). The reorganization required a standardized system of accounting. One offspring was the creation of ILSA of student learning outcomes. This development was created using a vision that if custom and law define what is educationally allowable within a nation, the educational systems beyond national boundaries suggest what is educationally possible (Foshay, Thorndike, Hotyat, Pidgeon, & Walker, 1962). This vision was used in introducing a first pilot study in mathematics and as such not only did it describe the origins of an emergent field but also foreshadow the exceptional growth of comparative assessment studies’ output to follow (Owens, 2013). Consequently, ILSA can be seen as a practice showing what is educationally possible.

The first organization formally instituted for this kind of activity is the *International Association for the Evaluation of Educational Achievement* (IEA). The founders viewed the world as a natural educational laboratory, where different school systems experiment to obtain optimal results in the education of youths. They assumed that if research could obtain evidence from different national education systems, the variability

would be sufficient to reveal important relationships, which would otherwise escape detection within a single education system (Pettersson, 2014).

The IEA is documented as attending a UNESCO meeting in Hamburg in 1961. In the documentation it is stated that during the week-long meeting there would be an opportunity to hold a discussion about what is known as the *Project on the Evaluation of Educational Attainment* (IEA). Even though IEA later was renamed, this is the first time that the acronym IEA appears. This mention did not signal the birth of the organization, because the organization had existed for quite some time. However, it does indicate that the IEA was an upcoming institution for achievement data and comparisons (Purves, 1987).

Already in 1955 an informal group of researchers had met at a joint meeting of the *International Bureau of Education* (IBE) and UNESCO to collaborate and learn about what they considered to be common educational problems. During the 1950s discussions were held about the American results in schools. Simultaneously, in Europe there was an increasing interest in comprehensive schools and how the secondary education could be improved. The desire for change on both sides of the Atlantic prompted cooperation beyond geographical borders. This, coupled with a desire to increase the educational measurement data with quantitative and explanatory data, led to a proposal being brought to UNESCO for an international study of intellectual functioning. The following arguments were put forward:

What has not heretofore been attempted, even on a limited basis, is a comparison that would take the school population near a terminal point, and involve many countries from the same general world culture [...] Such an effort, however, would have advantages: the results could be examined with one's mind on the fact that they arose from many apparently different conceptions of the nature and meaning of education; since the students were near the end of their formal education one might take the test responses as representing the outcome of the educational system as a whole, rather than catching a student in mid-career before the curriculum had been completed. (Foshay et al., 1962, p.7)

The purpose was said to be to make inferences about intellectual functioning from multiple-choice items, to test the feasibility of large-scale assessments, and also to be exploratory (Foshay et al., 1962). Here the study differs from other comparative studies in that it seeks to introduce an empirical approach into the methodology of comparative education, a field that is said to have initially relied on cultural analysis (Foshay et al., 1962). UNESCO agreed to support the international costs of such a study. The first meeting to discuss the study took place in Hamburg in the early summer of 1959 (Purves, 1987). Immediately after the meeting in Hamburg the group met again in London. At these meetings time was spent on discussing issues like sampling, test construction, questionnaire items and timetables.

The group embarked on the task with great enthusiasm and managed a pilot study (beginning in June 1959 and ending in June 1961) in which the researchers concluded that cross-national comparisons of educational performance could be made with comparable results (Foshay et al., 1962). Such findings were startling at the time, but even more important was the clear sense that a group of researchers from different cultures and educational systems could agree on a common approach to testing and evaluation (Purves, 1987). The original aim to study intellectual functioning seems to have been altered to include a much more sharply defined curriculum base in the test items. In this David Walker (1962) contributed with the phrase "*opportunity to learn*," which became one of the important items of study in the following IEA projects even though Walker's analysis in the pilot study found that individual ability accounted for more of the explained variance in the successful completion of an item than the teacher's emphasis in class (Walker, 1962).

Meetings about the joint assessment project were held in Hamburg during 1960 and 1961. In late 1960 there were discussions about doing another study of students' achievements, and mathematics was the subject proposed. In 1961, researchers from twelve countries met to discuss the pilot study that conducted assessments in mathematics, reading comprehension, geography, science and non-verbal ability. The study was considered a success and plans for another study in mathematics took shape. From the outset it was agreed that the project should be a co-operative enterprise. The major purpose of the inquiry was to measure achievement in mathematics and to relate that achievement to the relevant factors in the home, school and society. In determining these factors the investigation had to rely on the findings of previous research. The project called

the First International Mathematic Study (FIMS) was said to represent an attempt to assess the efficiency or productivity of different educational systems and practices (Bloom, 1969). The final results of FIMS were presented in a publication by Husre (1967). In addition to the main study, various reports were published (e.g. Keeves, 1968; Pidgeon, 1967; Kuusinen, 1967; Hultin 1968). In the study it became evident that there is a difference between how a subject actually is taught in the classroom and how it is described in the curriculum and that this is a good predictor of the differences in student performance. FIMS also showed that there is a lack of equity between different groups of students on how they perform. After this study the IEA performed a variety of studies on different subjects, time spans and periodicity. Table 3 follows a presentation of the various IEA studies:

Table 3. Presentation of the IEA-studies

| The name of the study and the acronym | Year of conducting the study | Year of publication |
|---|-------------------------------------|----------------------------|
| The Twelve-Country Study (The Pilot Study) | 1959-1960 | 1962 |
| First International Mathematics Study, FIMS | 1964 | 1967 |
| The Six-Subject Study: Science First International Science Study, FISS | 1970-1971 | 1973-1976 |
| The Six-Subject Study: Reading Comprehension | 1970-1971 | 1973-1976 |
| The Six-Subject Study: Literature Education | 1970-1971 | 1973-1976 |
| The Six-Subject Study: English as a Foreign Language | 1970-1971 | 1975-1976 |
| The Six-Subject Study: French as a Foreign Language | 1970-1971 | 1975-1976 |
| The Six-Subject Study: Civic Education | 1970-1971 | 1974-1976 |
| Second International Mathematics Study, SIMS | 1980-1982 | 1987-1993 |
| Classroom Environment Study | 1981-1983 | 1989 |
| Second International Science Study, SISS | 1983-1984 | 1988-1992 |
| Written Composition Study | 1985 | 1988-1992 |
| Reading Literacy Study | 1990-1991 | 1992-2002 |
| Computer in Education/Information Technology, COMPED | Two rounds: 1989, 1992 | 1991-1996 |
| Language Education Study, LES | 1995 | 1996 |
| Third International Mathematics and Science Study, TIMSS 1995 | 1994-1995 | 1995-2002 |
| Second Information Technology in Education Study Module 1, SITES-M1 | 1998-1999 | 2001 |
| Third International Mathematics and Study Repeat, TIMSS-R 1999 | 1998-1999 | 2000-2001 |

| The name of the study and the acronym | Year of conducting the study | Year of publication |
|---|---|----------------------------|
| TIMSS 1999 Video Study | 1998-2000 | 2003-2011 |
| Civic Education Study, CIVED | Two rounds: 1996-1997, 1999-2000 | 1999-2004 |
| Second Information Technology in Education Study Module 2, SITES-M2 | 2000-2001 | 2003 |
| Progress in International Reading Literacy Study, PIRLS 2001 | 2001 | 2001-2007 |
| Pre-Primary Education Project, PPP | Three rounds: 1986-1994, 1989-1993, 1993-2003 | 1994-2007 |
| Trends in Mathematics and Science Study, TIMSS 2003 | 2001-2003 | 2003-2005 |
| Second Information Technology in Education Study, SITES 2006 | 2005-2006 | 2008-2009 |
| Progress in International Reading Literacy Study, PIRLS 2006 | 2005-2006 | 2006-2007 |
| Trends in Mathematics and Science Study, TIMSS 2007 | 2006-2007 | 2005-2009 |
| TIMSS Advanced 2008 | 2007-2008 | 2006-2009 |
| Teacher Education and Development Study in Mathematics, TEDS-M 2008 | 2007-2008 | 2011 |
| International Civic and Citizenship Education Study 2009, ICCS 2009 | 2007-2008 | 2011 |
| Progress in International Reading Literacy Study 2011, PIRLS 2011 | 2010-2011 | 2012 |
| Trends in International Mathematics and Science Study 2011, TIMSS 2011 | 2010-2011 | 2012 |
| International Computer and Information Literacy Study, ICILS 2013 | 2012-2013 | 2014 |
| Trends in International Mathematics and Science Study 2015, TIMSS 2015 | Preparatory meeting 2013- | 2016 |
| Trends in International Mathematics and Science Study Advanced, TIMSS Advanced 2015 | Preparatory meeting 2013- | 2016 |
| Progress in International Reading Literacy Study 2016, PIRLS 2016 | Preparatory meeting 2013- | 2017 |
| International Civic and Citizenship Education Study 2016, ICCS 2016 | Preparatory meeting 2013- | 2017 |

| The name of the study and the acronym | Year of conducting the study | Year of publication |
|---------------------------------------|------------------------------|---------------------|
| Early Childhood Education Study, ECES | Preparatory meeting 2013- | 2018 |

Table 4. Indeed, it was the IEA studies that lead to many assessments being undertaken in various countries. The Programme for International Student Assessment (PISA) study, a project of the Organization for Economic Cooperation and Development (OECD), was similar to the IEA studies in many respects. Although OECD primarily has been concerned with economic policy, education has become increasingly important due to the fact that over the last 40 years education has been reframed to include economic competitiveness in an economic discourse related to human capital and “knowledge economy” (Pettersson, 2008). Through statistics, reports and studies OECD has activated a “common sense” in political decision-making by saying that scientific “proofs” are indisputable (Martens, 2007). Presentation of the OECD-studies

| The name of the study and the acronym | Year of conducting the study | Year of publication |
|---|---|---------------------|
| International Adult Literacy Survey, IALS | Three different test rounds: 1994, 1996, 1998 | |
| Adult Literacy and Lifeskills Survey , ALLS | 2003 | 2003 |
| Programme for International Student Assessment, PISA 2000 | 2000 | 2001 |
| Programme for International Student Assessment, PISA 2003 | 2003 | 2004 |
| Programme for International Student Assessment, PISA 2006 | 2006 | 2007 |
| Teaching and Learning International Survey, TALIS 2008 | 2008 | 2008 |
| Programme for International Student Assessment, PISA 2009 | 2009 | 2010 |
| Programme for International Student Assessment PISA 2012 | 2011 | 2012 |
| Programme for International Student Assessment PISA 2015 | Planned for 2015 | Planned for 2015 |
| Programme for the International Assessment of Adult Competencies, PIAAC | 2011 | 2013 |
| Teaching and Learning International Survey, TALIS 2008 | 2008 | 2009-2010 |
| Teaching and Learning International Survey, TALIS 2013 | 2013 | |

Martens (2007) argues that OECD's greatest impact can be seen in its agenda with indicators and its role in constructing a global policy field of governance by comparison (cf. Grek, 2009). Novoa and Lord (2002) state that comparisons like this may not be regarded as a method, but can in fact be seen as policy. The policy is driven by an expert discourse that, by means of comparative strategies, tends to impose natural or common sense answers in national settings (cf. Pettersson, 2008). While OECD serves national policymakers well with a comparable discourse in terms of statistics, it also provides them with a global policy lexicon concerning what education is and ought to be (cf. Pettersson, 2014). One way of creating this global policy lexicon is to look at what is measured in PISA. PISA provides comparisons of the competencies of 15-year-olds that are relevant to everyday adult life, rather than simply evaluating knowledge based on curriculum (OECD, 2001). It is also said that assessments that test curriculum only offer a measure of internal efficiency, and cannot reveal how schools prepare students for adult life (OECD, 2001). As such, PISA can be seen as a platform for policy construction, mediation and diffusion at national, international and even global level (Rizvi & Lingard, 2006). It can be said that on the one hand PISA shapes an international discourse and on the other PISA is shaped by the discourse.

PISA assessments have been conducted several times. In every assessment students' knowledge in reading, mathematics and scientific literacy is tested, together with interests and backgrounds. In addition, innovative domains are also assessed, for example what is called collaborative problem solving, and there are also plans for incorporation of what is called global competency from the 2018 assessment onwards. The emphasis on "real-life" circumstances and the capacity to enter the labour market with the relevant skills has been said to shift PISA's focus away from less explicit educational aims that are more complicated to measure (Grek, 2009). PISA also easily connects to the idea of the self-governance of active subjects, which expands governance into a system of individual self-regulation (Ball, 1998). Even though PISA is both constructed and operates under a clear policy framework that is designed to improve future results, it is not just a testing regime. PISA should also be seen in light of its ability to improve and attract economic and human capital investments. For policymakers, PISA is therefore a two-sided coin in that it tests outcomes and attracts economic investment. In view of this, PISA can be said to have two functions – economic and educational – in international policy discourse (Pettersson, 2008). As these two aspects are interwoven and strengthen each other, they can hardly be analyzed separately. Besides PISA, OECD has also staged and presented various other studies. Table 4 is presenting the various studies published by the OECD.

Through the 1980s and 1990s, ILSA proliferated in type and design. Above in Table 4 it is illuminated how IEA expanded their surveys from mathematics and science into reading, pre-primary education, classroom environment, second language acquisition, technology and civics. However, while PISA imitated the 1959-61 study by the IEA in the intent to evaluate student performance close to the end of schooling, the objective of the assessment had evolved distinctly from curriculum-based learning to a new concept of literacy. Literacy was employed by PISA to signify a mastery of broad concepts, applicable to life beyond the classroom (OECD, 1999).

Since the introduction of PISA, discussion has been flourishing on how to describe and distinguish between surveys performed by the IEA and the OECD. One way to explain these differences is to describe IEA as more research-oriented and OECD as more policy-oriented (Pettersson, 2008). In a comparison between 22 participant countries in the 2003 PISA and TIMSS cycle it is concluded that despite differences of curricular and literacy intent, country results are comparable (Wu, 2009). Comparable is also the methodological construction and the applications for further research emanating from PISA and TIMSS publications (Hutchinson & Schagen, 2007). Owens (2013) concludes in this debate that the aggregated "league table" results of the ILSA have caused more dismay than the psychometric considerations underpinning the tests. Connected to this development several international, regional and national agencies have tried to develop systematic studies of various aspects of education and schooling, mostly intended to evaluate national educational quality.

Regional Assessments

Concurrently, another kind of learning assessment also developed – regional assessments. In these assessments instruments were developed to assess and compare learning outcomes among nationally representative samples

of students within a particular region. For Europe the most evident organization conducting these kinds of studies is the EU with e.g. its knowledge assessment ESLC (European Survey of Language Competences). In collaboration with the UNESCO two other regional organizations, one in Latin-America and one in Africa, conduct regional learning assessments: the Latin American Laboratory for the Assessment of the Quality of Education (LLECA) and the Southern and Eastern African Consortium for Monitoring Educational Quality (SACMEQ). Besides, the Conférence des Ministres de l'Éducation des Pays Ayant le Français en Partage (COMFEMEN) conducting the assessment PASEC (Programme d'Analyse des Systèmes Éducatifs de la CONFEMEN) in Francophone countries conducts regional learning assessments. All these regional bodies spurred and developed important regional assessment exercises with the support of international, regional and national experts, as well as with the help of national and international funding sources.

These regional organizations are very much interconnected to both the methodology in use in the ILSA as well as the personnel involved in developing and conducting the international tests. An evident example of this is when SACMEQ first was developed it was through a meeting between the Zimbabwe's Minister for Education and the Director of UNESCO's *International Institute of Educational Planning* (IIEP) where they agreed that a major research and training project called the *Indicators of the Quality of Education Study* should be undertaken in Zimbabwe during the 1990s in order to (a) assess the quality of education provided by primary schools, (b) involve the staff of the Ministry's Planning Unit in integrated research and training activities, and (c) provide meaningful advice related to policy concerns expressed by senior Ministry decision-makers. The project resulted in a research report written by Kenneth N. Ross and Neville Postlethwaite in 1991. Both Ross and especially Postlethwaite were prominent contributors within the IEA where they, among other things, had cooperated in the late 1980s in the *Reading Literacy Study* (Postlethwaite & Ross, 1992). Even though the project was an internal one it leads to that the IIEP initiated follow-up training workshops for educational planners and researchers from Zimbabwe and also for several nearby countries. In the 1990s educational researchers and planners that had participated in these workshops commenced a dialogue with their respective Ministries of Education and the IIEP staff on the training needs required in order to expand and strengthen the capacity of their Education Planning Units to monitor and evaluate the quality of their education systems.

This dialogue eventually resulted in the establishment of an association of Ministries of Education under the acronym SACMEQ. The organization took on the challenge to develop cross-national cooperative activity with the help of IIEP and other UNESCO officials, and after the first report in 1995 (authored by teams of educational planners and researchers from ten countries) the organization and their cross-national ambitions were very well appreciated. In 1995 the SACMEQ Consortium was officially launched and given long-term support through the financial assistance of the Government of the Netherlands. Today the organization comprises nations like Angola, Botswana, Kenya, Lesotho, Malawi, Mauritius, Mozambique, Namibia, Seychelles, South Africa, Swaziland, Tanzania (both the mainland and Zanzibar), Uganda, Zambia and Zimbabwe. Within these nations the mission is to undertake integrated research and training activities that will expand opportunities for educational planners and researchers. This will be done through (a) receiving training in the technical skills required to monitor, evaluate and compare the general conditions of schooling and the quality of basic education, (b) generating information that can be used by decision-makers to plan the quality of education and finally, (c) utilizing innovative information dissemination approaches and a range of policy dialogue activities in order to ensure that the results are debated, discussed and understood by stakeholders and then used as the basis for policy and practice. SACMEQ have until today conducted four major studies (I-IV). The first study conducted in 1995 and completed in 1999 concerned educational inputs in schools, benchmark standards for educational provision, equity in the allocation of educational resources, and the reading literacy performance of Grade 6 learners. The data included information from around 20 000 learners, 3000 teachers and 1000 school principals in seven nations.

Interconnectedness to ILSA is explicit because one of the writers of the technical report is no other than the current director of the OECD's PISA-studies, Andreas Schleicher (Schleicher & Saito, 1995). The report is written under the auspices of the IIEP. SACMEQ II was conducted in 2000 and completed in 2004 among fourteen countries; this time the assessment was enlarged to also comprise mathematics. Twice as many learners are tested (40 000) as well as 4000 teachers and 2000 school principals. The assessment is reported in national reports, but it is evident that all the technical reports have been published by the IIEP. A third round of

SACMEQ was conducted in 2006 and completed during 2011. This time comprising fifteen countries focusing on literacy and mathematics, and it also expanded to cover learners' levels of basic knowledge about HIV and AIDS. This time 61 000 learners, 8000 teachers and 2800 school principals were involved. The assessment resulted in far more publications. The two first assessments resulted in 7 National Reports and 5 Technical Reports (SACMEQ I) and 14 National Reports and 4 Technical Reports (SAQMEQ II). With SACMEQ III the publication rate and the organization of the reports change character, resulting in 54 Policy Briefs, 12 National Reports, 2 Working Documents, 8 Working Papers and 8 Policy Issue Series. The SACMEQ IV is scheduled for implementation during 2012-2014 with a new innovation that would enable participating countries to track changes over time, to make valid comparisons with similar education systems and to report on progress made towards the attainment of education for all.

Table 5. Regional assessments

| The organization performing the assessment | Year of conducting the study | Year of publication | Name and acronym of the assessment |
|--|-------------------------------------|----------------------------|--|
| EU | 2011 | 2011 | European Survey of Language Competences, ESLC |
| LLECA (the Latin American Laboratory for the Assessment of the Quality of Education) and UNESCO | 1997 | | The First International Comparative Study in Language, Mathematics, and Associated Factors in the Third and Fourth Grades of Primary Education |
| LLECA | 2006 | | the Second Regional Comparative and Explanatory Study, SERCE |
| SACMEQ (the Southern and Eastern African Consortium for Monitoring Educational Quality) and UNESCO | 1995 | 1999 | SACMEQ I |
| SACMEQ | 2000 | 2004 | SACMEQ II |
| SACMEQ | 2006 | 2011 | SACMEQ III |
| SACMEQ | 2012- | | SACMEQ IV |
| CONFEMEN (Conférence des Ministres de l'Éducation des Pays Ayant le Français en Partage) | 2012- | | PASEC |

Latin American countries have only sporadically participated in ILSA, but in the two regional LLECE tests their desires to participate have been much more evident. Starting in the 1990s and aided by the UNESCO Latin American countries started to develop tests in mathematics and reading skills that were to be implemented across the region. The overall aims of the project were to (a) increase pertinence and reduce the cultural bias of its own and other international studies, (b) conduct international comparative and explanatory analyses of students achievements, (c) identify input, context and process variables dealing with social conditions of the students and school, family, curricula, and other school, teacher and student factors that impact on student achievement, and (d) provide useful information for the formulation and execution of educational policy in the countries of the region. In 1997 the LLECE carried out the *First International*

Comparative Study in Language, Mathematics, and Associated Factors in the Third and Fourth Grades of Primary Education (Hanushek & Woessmann, 2012), especially designed for testing educational achievement in Latin American countries. The assessment tested the performance in mathematics and reading of representative samples of students in each participating country in primary schools, grades third and fourth. In 2006 another test, said to be especially design for the Latin American countries, was launched in the region with the support of the Latin American bureau of the UNESCO called the *Second Regional Comparative and Explanatory Study*, with the acronym SERCE. The second study also tested the performances in mathematics and reading with representative samples, but this time the cohort comprises students from third and six grades.

Another regional organization with a different history is CONFEMEN founded in 1960 by member states within the francophone world, bringing policymakers together for discussing education. In 1991 at a meeting it was stated that there was a need for bringing together all the quantitative and qualitative aspects of the educational systems studied and for identifying the most effective educational strategies. In the last two decades PASEC's (the acronym for the assessment staged) mission to evaluate performances of education systems have resulted in 35 national assessments in more than twenty countries in Africa and Asia. Since 2012 PASEC also implements international comparative assessments to, as they say, better meet the needs of the country. These international assessments are not international in the orthodox meaning, instead international in the PASEC interpretation means that the assessments comprises several of the member states within the francophone cooperation. The main objectives are (a) to measure student performance and identify factors of effectiveness and equity for basic education, (b) to provide national policy indicators that allow member countries to compare in space and time, (c) to continue the development of an internal and permanent evaluation of their educational system capacity, and finally (d) to disseminate international assessment results to contribute to the discussion and debate on the determinants of the quality of education. PASEC measures the achievements in French (and/or national language if that is the language of instruction) and mathematics. Simultaneously a collection of contextual, institutional, social, economic and cultural data is taking place. In explaining the academic performances of the students these data allow separate effects among those due to the educational context.

Also the regional assessment within the EU has to be considered when discussing the evolution of these kinds of tests. The *European Commission* describes the purpose of their assessment ESLC as providing participating countries with comparable data on foreign language competence and knowledge about good practice in language learning. The test is also intended to provide indicators for measuring progress towards the objectives of improving foreign language learning. The survey tested the two most widely taught European languages in each country among students in their final year of lower secondary education. The survey assessed students' proficiency in listening, reading and writing. In addition to the test the students were asked to fill in a questionnaire about their language learning and other background factors, this for providing data on how demographic, social, economic and educational variables affect language proficiency. Evident is the influence from the ILSA using both tests and questionnaires for addressing the issues emphasized.

Table 5 is summarizing the assessments mentioned.

National Assessments

Over the years the prominence of ILSA, especially the ones conducted by the IEA, and with support from different IGOs and NGOs for evaluations of student learning, technically and financially, contributed to the spread of a new form of assessments – national assessments (Kamens & Benavot, 2011). These non-standardized assessments of cognitive achievement, but sometimes also attitudes and behavior, were conducted under the auspices of the countries themselves, but not unusually with support from different NGOs and IGOs or international donor agencies. Also the help with educating and disseminating assessment personnel within the IEA is important (Pettersson, 2008). These assessments are not to be considered as absolutely new ones. Different national assessments have been conducted more or less regularly in history, but they have virtually exploded during the last 20 years, especially considering the developing world (Kamens & Benavot, 2011). This is due to that performance in schools is within the global educational “*common sense*” increasingly judged on the basis of effective student learning outcomes. In this information and data is seen as critical for

knowing whether the school system is delivering good performance, and providing feedback is the channel through which performance can be improved. Increasingly, countries are therefore developing a range of tools and techniques for evaluation and assessment in school systems as part of their efforts to improve their students' outcomes (Rosenkvist, 2010). At first glance, the make-up of these national assessments may appear somewhat identical, but in fact there are often substantial differences between countries in implementation, use and design. The differences mostly arise from the fact that assessments are political phenomenon reflecting the agenda, tension, institutional norms and the very nature of power relations between different internal political actors (Kellaghan et al., 2009).

A national assessment is a survey of schools and students that is designed to provide evidence about the levels of student achievement in identified curriculum areas for a whole education system or for a clearly defined part of an education system. The main focus is normally to describe and evaluate the quality of student learning outcomes that have been produced by schools. During the 1970s and 1980s a number of industrialized countries established national assessments on a more regular basis. In the early 1990s many more countries became interested in conducting regular assessments. This heightened awareness about the importance was encouraged by the declaration of the *World Conference on Education for All*, held in Jomtien, Thailand, in 1990 which stated that providing students with access to education was meaningful only if the students actually acquired useful knowledge. The *Dakar Framework for Action* in 2000 reinforced this message and also stressed the importance of having a clear definition and accurate assessment of learning outcomes. These declarations were especially important for developing countries in staging national assessments (Kamens & Benavot, 2011).

Similarities are evident in the subject matter studied in national assessments across the world. Mostly reading comprehension and mathematics are assessed. Some have included science, based on either general knowledge or focused on topics in biology, chemistry and physics. A few national assessments have included art, music or civic education. At the secondary level some interest has been shown in the skills of reading, writing, listening and speaking a foreign language (Postlethwaite & Kellaghan, 2008).

Within the report *Strong Foundations: Early Childhood Care and Education* (Encinas-Martin, 2006) a preliminary overview of national assessment activities is provided. In a UNESCO follow-up study, Aaron Benavot and Erin Tanner (2007) analyze this statistics more thoroughly, concluding that national learning assessments have become a common feature of national education systems around the world. The total number of countries performing at least one national assessment has risen steadily over time, from 65 countries (1995-1999) to 111 countries (2000-2006). The report is written within UNESCO, and as an answer to why the assessments have become more common they state that it is due to the declaration in Dakar. The statement might be true but as an answer to why national as well as international and regional assessments have increased that explanation is rather shallow. Instead it is probably more accurate to attach the ideologies behind the Dakar declaration as the producer of this increasing practice of assessments. This ideology contains both comparisons and the use of data as its prerequisites.

The subjects assessed in the national assessments are said to be predominantly curriculum-based and subject-oriented. Broken down into five major subject categories the UNESCO report (Benavot & Tanner, 2007) shows that mathematics and language are by far the most prominent subjects evaluated with more or less every country staging national assessments assess these subjects. This not only reflects the strong emphasis on mathematics and language in national curricula worldwide, it also shows the centrality of these subjects in international assessments and educational discourse. Half of the countries staging national assessments assess learning outcomes in science and almost two-fifths assess learning in the social science. One-fifths in foreign languages and nearly the same is true for other subjects like art, physical education, problem solving, life skills, visual literacy, colouring, cognitive behavior and music. It is also shown that there has been little change over time in the prevalence of assessed subjects, except foreign language which have gained prominence and social science which have lost the same.

From research to policy and back: assessing impact

As showed, the origins of ILSA and other regional and national assessments can be traced back to the early investigations undertaken by the IEA with the emphasis on developing a collaborative effort directed towards

understanding the relationships among the inputs to education and its outcomes. As such it was said that the early ILSA could result in revelations and produce powerful insights for educational reforms and improvements.

The expansion of ILSA is well documented (cf. Kamens, 2009), but questions continue to be asked about the extent to which ILSA has an impact on the education systems of the participating countries. The question most often asked is whether ILSA contributes to educational reforms and improvements in respective countries. The question is understandable considering that most organizations behind the ILSA have more or less clearly articulated aims within this strive. The categorization and the description of impact factors stemming from ILSA have been taken from Hans Wagemaker (2014) in *Handbook of International Large-Scale Assessment* as departure for describing the area. The first seven categories have been collected from his work, but the last – *Research Generating* – is new. This is because of this report's aim to say something about what kind of research ILSA is leading to. In this report we are not presenting the results but discussing to what amount ILSA have led to an increasing interest among academics to use and present results stemming from the ILSA. Even though Wagemaker is the main source for this section, some references and content have been added to better describe how researchers discuss ILSA impact.

There are evidently a lot of challenges in assessing impact, and numerous analysts and researchers have tried to identify the challenges in judging the impact of ILSA. When reflecting on why researchers within the field of evaluation were having little impact on policy matters, Caplan (1979) noted that the two communities – theorists and policymakers – were living in different worlds, divided by different values, reward systems and often different languages and probably also act within different styles of reasoning (Hacking, 1992a). As an effect researchers trying to judge impact often end up in a rather pessimistic conclusion as to its efficiency (Burkhardt & Schoenfeld, 2003). Wagemaker (2014) concludes in that:

[A] Common conclusion is that stakeholders need to recognize that major policy initiatives or reforms are more likely to result from a wide variety of inputs and influences rather than from a single piece of research. Research is also more likely to provide a heuristic for policy intervention or development rather than being directly linked, in a simple linear fashion, to a particular policy intervention. This outcome is due not only to the competing pressures of interests, ideologies, other information, and institutional constraints, but also to the fact that policies take shape over time through the actions of many policymakers. (Wagemaker, 2014, pp.12-13)

Even though these challenges in judging impact exist, we will now with the help of the Wagemaker (2014) review try to introduce different ways to discuss and measure ILSA impact on policy and research.

Growth of ILSA as a measure of impact

Wagemaker (2014) states that it might seem natural to state that the growth of ILSA and demands for participating in ILSA is evidence enough for stating the impact of ILSA. This growth has been described as an “*explosive growth*” (Kamens, 2009) of ILSA and national testing, especially among the developed countries because they are attributed to an emerging consensus about not only the legitimacy but also the necessity of assessments. As such ILSA is now regarded as a regular feature of the educational landscape even though they are a relatively recent phenomenon that has evolved in line with increasing demands on accountability (Wagemaker, 2014).

As showed the history of ILSA dates back to the early 1960s, but Wagemaker exemplifies a newer development. The report by the U.S. *Commission on Excellence in Education* (Gardner, 1983) in many aspects marked the beginning of a new era for ILSA and their expansions. The report expressed concerns about educational performance in both the US and in many OECD countries. When introduced, a consensus emerged that national goals for education were in need. These goals would focus on improving the performance of elementary and secondary students and include expressed expectations on the American students, but it was also stated that the performances would exceed student performances in other countries (Ambach, 2011), and for measuring these goals international references were in need, and as such ILSA grow in importance. Even though this realization mainly concerned the American context, a simultaneously evolution occurred within

the OECD nations that the education systems need to operate in a supra-national space, responding to the demands to educate citizens capable to compete in a competitive, rapidly changing globalized social, economic and political world (Wagemaker, 2014). The report, which had a high degree of visibility by the time it was published, also marked a clear move away from expanding education systems in terms of enrollment towards a more systematic focus on national monitoring and questions concerning educational excellence, equity and efficiency (Tuijnman & Postlethwaite, 1994). This moving away from enrollment as the prime concern is mainly affecting developed countries. Today, in times of *Education for All and the Millennium Development Goals*, enrollment is basically only of great concern for developing countries and not so much for developed countries. As such, participating in ILSA can mark a transition for developing countries from enrollment as the basic focus more into focusing on educational quality (Wagemaker, 2014).

The changing in focus, from enrollment to quality, can be illuminated with the help of the growing numbers of nations participating in ILSA, but also with the help of nations participating or staging regional and national assessments. The manifestation of an increasing interest for participating in ILSA can be shown with the help of IEA's studies of mathematics and science performance. Since IEA's first study conducted in the early 1960s the association has carried out more than 26 major studies examining differences in student performance in and across nations (Mullis, Martin, Gonzales et al., 2004). The participating nations started among 12 countries but the nations involved have continually grown. When it was time for staging the *Second International Mathematics Study* (SIMS) in the early 1980s the participating nations had grown to 20 and the enrollment of nations continued to grow. TIMSS 1999 was conducted in 40 educational systems; TIMSS 2007 was conducted in 66 educational systems. By the time of TIMSS 2011, the involved educational systems had increased to 79 and also including a number of benchmarking US states and other subnational systems such as Dubai (one of the United Arab Emirates nations). The same pattern of growth can be seen when studying PISA launched in the late 1990s among 43 OECD countries, but today the assessment attracts participants and interest far beyond the original OECD membership (OECD, 2006) which can be visualized with the help of involvement in the PISA 2012 round where 68 educational systems are involved (also including subnational entities like Macao and Dubai). Also evident when studying the participation in both IEA and OECD assessments is the expansion beyond developed nations to also including less developed countries as well as a growth of subnational entities like US states or Dubai and Macao choosing to benchmark their educational performance against standards set by the rest of the world.

Today, the work of IEA, especially evident in its TIMSS and *PIRLS (Progress in International Reading Literacy Study)* assessments, along with OECD's PISA study are truly worldwide and of international concern. The growing number of ILSA has also precipitated other cross-national assessment initiatives, such as the regional assessments PASEC, LLECE and SACMEQ. All these assessments have a history dating back to the 1990s when initiatives were taken among smaller groups of countries with narrow scopes, but since then both participating countries and the scope of investigation have considerably grown. Another contemporary trend is that both IEA and OECD have gone beyond their traditional scope of investigating students' performance. In 2010 IEA reported on their *Teacher Education Development Study-Mathematics* (TEDS-M). The study is a major investigation of the pre-service training of mathematics teachers in primary and lower-secondary schools in 17 countries (Tatto et al., 2012). The same has become true for OECD when conducting studies like *Program for the International Assessment of Adult Competencies* (PIAAC) conducted across 25 OECD member countries on the adult population's (aged 15-65) literacy and numeracy skills and their *Teaching and Learning International Survey* (TALIS) that focuses on issues surrounding the teaching and learning environment in schools and teachers' working conditions. Studies like these have taken ILSA outside their traditional domain of primary and secondary classrooms, and in doing so expand their sphere of influence among policymakers and researchers (Wagemaker, 2014).

Finally, Wagemaker (2014) states that the recognition of a more globalized educational community also shapes a shift from a more tactical approach for participating in ILSA to a more strategic, investment-based participation on a more regular basis. As a result many countries choose to participate in several different ILSA dependent on the different perspectives the ILSA provide. This is especially evident when comparing participation in OECD and IEA studies in which it becomes evident that if countries choose to participate in a OECD study they often also participate in IEA studies and vice versa. An example of this is that 76 percent of

17 developing countries participating in PISA 2006 also previously have participated in one or more IEA studies.

Discourse as a measure of impact

Wagemaker (2014) is writing that there is a lot of research investigating the impact of ILSA in shaping discourses about education, especially evident in terms of discussions about educational quality. Wagemaker (2014, p.20) makes an Internet search on different ILSA references and the result is speaking for itself, TIMSS is referenced around 5.21 million listings, PIRLS is listed for some 3.25 million listings and PISA gives around 4.37 million. The listings are related to numerous education systems, professional organizations, governmental and nongovernmental bodies, and interested professionals. When we make the same search but using Google.se “PIRLS” is listed around 524 000 times and “*Progress in International Reading Literacy Study*” is listed around 1.2 million times, “TIMSS” is listed around 936 000 times and “*Trends in International Mathematics and Science Study*” around 33.6 million times, and finally, “PISA” gives 13.1 million hits and so does the “*Programme for International Student Assessment.*” The extent to which information resulting from different ILSA has entered public domains can be seen as evidence on the importance of the studies both in media and the general public, but also in terms of the political debate the release of ILSA has instigated in various nations.

One example uplifted by Wagemaker (2014) on how the results from ILSA is used in the political debate is when Alan Greenspan, then the Chairman of the US Federal Reserve, in 2004 used the information from TIMSS in his address to the *Committee on Education and the Workforce of the US House of Representatives*. In his address (Greenspan, 2004) he expressed concern about the educational standard among the US population and the negative impact that this underachievement would continue to have on the US national economy. What becomes evident in this address is that the results in ILSA can lead to internal debates about national economy in a comparative globalized world. Another example is when the former Swedish Minister of Education, Jan Björklund, used results from ILSA in arguing for reformation of the educational system. Political turmoil stemming from the presentation of results within different ILSA is therefore an example on how ILSA can impact national policy. Results have been described as both “*TIMSS-shocks*” (Lehmann, 2011) and “*PISA-shocks*” (Ertl, 2006).

Changes in educational policy as a measure of impact

Participating in ILSA is in many countries directed towards improving and understanding educational provisions. This has traditionally focused on input in education. Quality has often been judged on the basis of these input measures. ILSA is more focused than many other investigations in attempting to capture these background (input) variables in order to better understand educational outcomes. Several of the international and regional assessments – such as TIMSS, SACMEQ, LLECE and PASEC – also try to capture information on factors relating to curriculum and instruction and their association with educational achievement. One reason for this interest is that these factors are deemed capable for policy interventions (Wagemaker, 2014). The national analysis of ILSA data can therefore be seen as influencing curricular and instructional reforms (e.g. Aggarwala, 2004; Elley, 2002; Mullis and Martin, 2007; Robitaille, 2000; Wagemaker, 2011). One factor that can be uplifted in this discussion is the fact that countries use TIMSS and PIRLS results as an integral means of informing on internal reforms and strategies (Wagemaker, 2014). A striking example of this is presented by Lehmann (2011) when saying that ILSA such as PISA, PIRLS and TIMSS is mandated by German law and that they are “*a necessary condition for policy making*” (a.a., p.422).

There has been several evaluations performed on the impact of IEA studies (some of them presented by Wagemaker, 2014 are e.g. Aggerwala, 2004; Lockheed, 2009; Robitaille, 2000; Schwippert, 2007; Schwippert & Lenkeit, 2012); all conclude that the studies are of great importance in discussing impact on national reforms and policy. This is also true for evaluations discussing PISA impact on national policy. One example of this is Breakspear (2012) stating that PISA played a key role in monitoring educational standards and issues related to equity and accountability. Normally questions about ILSA impact are addressed with references to TIMSS, PIRLS and PISA. However, other studies such as the IEA studies *Second Information Technology in Education*

Study Module 1 (SITES-M1) and the *Civic Education Study (CIVED)* have played a crucial role in the educational debate (Wagemaker, 2014).

Curriculum changes as a measure of impact

Wagemaker (2014) is making an argument that some of the ILSA have a more outspoken focus on curriculum. Especially the assessments managed by IEA (e.g. TIMSS and PIRLS) and the regional SACMEQ assessments are designed not only for describing differences in achievements within and across countries but also addressing questions as to why these differences exist. One of the more outspoken focuses in these assessments is related to curriculum and teaching practice. Wagemaker (2014) states that from the curriculum changing perspective even among countries that achieve well on the ILSA these assessments appear to have a substantial influence on the ways in which countries manage the process of educational reforms and improvements. After this statement he gives a variety of examples about countries using the results in the different TIMSS assessments and other ILSA for changes in both curriculum and teaching practice. What is noteworthy in these descriptions is that countries seem to use the results in different ILSA for legitimating and processing different national reform initiatives. As such, it might be said that ILSA are used as a content and provider in describing the national situation and as such have become a natural part in national reform processes. This description has a lot in common with how PISA is used internally in the policymaking of Swedish education as a tool for steering (Pettersson, 2008).

Changes in teaching as a measure of impact

ILSA also have an impact on the preparation of teachers (Wagemaker, 2014). According to Wagemaker there are numerous examples of countries using ILSA findings for changes in their teaching and instruction policy and practice. These changes have in different countries taken various forms – such as changes in teachers' training, teacher-training materials and toolkits for improvements, teachers' handbooks and so forth. Even though Wagemaker concludes with this, we can in our systemic review note that the results based on empirical knowledge in this field is rather weak, as will be presented later on in this text.

Capacity building and research endeavors as a measure of impact

Wagemaker (2014) introduces this section by saying that several of the ILSA need to be understood as both process and product. By this he means that the agencies responsible for developing and coordinating ILSA spend a lot of time and effort to ensure that the participants benefit from their participation not only in terms of outcomes but also in terms of the training they receive. The impact in nations of this training takes several forms.

One of the most evident forms of impact is that the training of regional coordinators and specialists in educational measurement and data analysis leads to adopting national assessments highly inspired by the methodology in the ILSA. The influence that ILSA have on educational capacity-building and research endeavors in general in participating countries can also be seen at the institutional level. One of the most striking examples of this is how the Australian Council for Educational Research (ACER) is transformed from a national endeavor into a large and highly respected organization (Ainley et al., 2011) with a lot of responsibilities for staging data analysis for different ILSA. The ILSA has as such had a substantial impact in the area of research and assessment methodology. This is in fact one of the motivating factors for developing countries to get involved in ILSA. In a study by Aggerwala (2004) TIMSS is especially uplifted as an impact factor for developing countries where it is concluded that TIMSS has led to meaningful educational reforms in curriculum, evaluation, assessment standards and tests and examinations (in this Egypt is especially emphasized as an example), reforms in curriculum, testing and methodologies have been institutionalized, growing awareness of the need for reforms in areas like teaching methodologies, teachers' training and licensing, a greater adaptation of new test formats and scoring methodologies, a larger recognition of a shifting emphasis from learning to application of knowledge and finally, a renewed awareness of the essential need of educational reforms for a knowledge-based society.

Another line of development stemming from the ILSA is that prior to involvement in ILSA a lot of nations have rather poor data on educational outcomes (e.g. Howie, 2011; Saito & Capele, 2010), but involvement in ILSA helped countries to develop local research capacity in assessments.

Today, ILSA are part of the educational assessment landscape in many countries and are often taken for granted. As such their evolution and expansion represents a significant development in research and measurement methodologies. A lot of the ILSA are originally based on methodologies developed within the *US National Assessment of Educational Progress* (NAEP) (Wagemaker, 2014). The process of extending this model for allowing comparisons of educational achievement both within and across a larger number of entities has posed several major challenges. One of the challenges is connected to the consequential validity of conducting these assessments. As a result the ILSA have to include the highest technical standards possible to ensure reliability and validity. ILSA are considered to be low-stakes assessments in terms of the potential impact on individuals; they are potentially high-risk enterprises with respect to the decisions countries make as a consequence of the findings within ILSA. A way to handle these difficulties is that most of the ILSA publish technical reports providing those who analyze or read the reports with the guidance necessary to appropriately interpret the information. In addition, workshops are given e.g. AERA often has a pre-conference workshop that utilizes ILSA data to provide technical training for those that will use the data for substantial work. A consequence of these difficulties with the high technical skills needed for interpreting and properly analyzing the data has created a specialized staff within governmental and non-governmental organizations and research and as such created a specific field for working with and interpreting ILSA data. As such a specific research, or semi-research, field has been created with highly specialized researchers and analysts. Even though a lot of the ILSA state that the educational reforms and improvements are in focus rather than the assessment or the test per se, a lot of the organizational bodies behind the ILSA creates imperative to ensure that the data gathered are readily accessible and used. Examples on this are that both the IEA and the OECD encourage and facilitate analysis of the ILSA data through targeted research conferences and accessible online data analysis tools. Training in data analysis is staged by several organizations, but to ensure even more that the data is correctly used some organizations like the IEA provide specific data analysis tools like the *International Database Analyzer*. Use of these tools has been further supported with the help of regular series of training seminars offered by sponsoring organizations e.g. *US National Center for Educational Statistics* and the World Bank. Another way to ensure community building with researchers is to collaborate and establish communities like IEA-ETS (*Educational Testing Service*, Princeton, New Jersey, USA) Research Institute (IERI). In providing training for basic and advanced analytical techniques related to ILSA they annually publish a monograph called *Issues and Methodologies in Large-Scale Assessments* to provide a forum for research related to different aspects of ILSA (Wagemaker, 2014). Connected to these organizational bodies is a biennial international research conference (IRC) that offers an academic forum for analysis, discussions and disseminations of data related to IEA studies. Also, a Springer Open Access Journal was introduced in 2013, called *Large Scale Assessment in Education*.⁴ The journal is a joint publication of the IEA and *Educational Testing Service* (ETS). The articles in the journal are said to contribute to the science of ILSA, help disseminate state-of-the-art information about empirical research using the databases of ILSA and make the results available to policymakers and researchers around the world.

Besides the contributions to methodology and technology offered by the ILSA they have also contributed to educational theory (e.g. Keeves, 2011). During the development of different ILSA they have attempted to gain better understanding of the impact of schooling on student achievement. A key contribution of these efforts has been to move discourses about quality of education away from the input side to the output side. In doing so, ILSA have made learning outcomes the central focus for assessments. This process has been accompanied by attempts to include the broader social context for learning, such as students' home-background experiences, school environment and instructional practice and how these factors are related to achievements.

⁴ <http://www.ierinstitute.org/largescaleassessmentjournal.html>

According to Wagemaker (2014) the most enduring feature of the early IEA works that have later on affected other more contemporary ILSA is the separation in intended, implemented and achieved curriculum. In this model the intended curriculum stands for the curriculum dictated by policy, the implemented curriculum stands for what is taught in the classrooms, and the achieved curriculum is what the students actually learned and can do with this knowledge. The model has come to become deeply imbedded in both the IEA studies and other ILSA.

Global and donor responses as a measure of impact

One of the more evident global trends in education today according to Wagemaker (2014) is the emphasis stemming from the *Millennium Development Goals and Education for All*, saying that the priority for national educational systems is to achieve universal primary education. There is a great difference in this emphasis in relation to ILSA in that the former is rather un-interested, at least explicitly even though the contrary has been stated elsewhere in this report, in the quality of education. Organizations like the World Bank have therefore been rather early advocates for a more focused debate heading for quality, and in this ILSA have been promoted as a way for better focus on quality. This focus can easily be seen in the World Bank publication *Learning for All: Investing in People's Knowledge and Skills to Promote Development* (World Bank, 2011a). Within this publication and other publications developed in donor or development organizations like the *US Agency for International Development* (USAID) and the UK's *Department for International Development* (DIFID) is their reliance on evidence from ILSA to support their arguments for greater emphasis on learning outcomes, but also for their lobbying efforts to promote more developing countries to participate in ILSA (Wagemaker, 2014). Another such emphasis on outcomes can be seen within the work of *Fast Track Initiative Countries headed by the Global Partnership for Education*.⁵

ILSA as a promoter for economic development has also attracted a variety of developmental economists. This has probably helped organizations performing ILSA to secure funding for their assessments. As a result we can today see that organizations like the World Bank and the *United Nations Development Program and the Inter-American Development Bank* all provide funding for staging ILSA. These organizations have also supported participation in ILSA for less-developed financially countries. When Wagemaker (2014) is discussing this area of impact he states that the influence of ILSA on policy and practice on development organizations such as the World Bank is nowhere more evident than in the establishing of *Arab Regional Agenda on Improving Education Quality* (ARAIEQ). The organization is established for addressing educational quality in the Middle East and North Africa region (World Bank, 2011b). In the proposal for the establishment of this organization it becomes evident that data from both PISA and TIMSS are heavily implicated for establishing a series of regional networks emphasizing on improving educational outcomes.

Research generating as a measure of impact

This might be said to be one of the overall focuses of this report, but nevertheless some insights might be handled on an overall level. First, it might be said that there is not much research made on the impact of ILSA in scientific work. Questions that might be raised in this discussion are questions dealing with what kind of research is generated, by which researcher, where do they publish their results and so forth.

One of the few meta-studies found on research generated from ILSA is made by Dominguez, Vieira, and Vidal (2012). In their specific case they are investigating PISA. In this they state that PISA reports have had a considerable impact on the scientific research if considering the results published in academic journals. The authors have found a pattern in publishing in academic journals that they state would be reasonable in this kind of fields of research. The interest in PISA is reflected primarily by university-based authors from the areas of education and the social sciences. They have also found that the articles over time have grown in numbers. The

⁵ www.educationfasttrack.org/

most analyzed topic was the same that is the core of PISA: students and their performance. This expected pattern means that researchers use PISA as a new but common source of information for their research. The authors also found that the impact on science from PISA seemed to be larger on the boards of journals than on individual researchers. This is because a lot of journals have presented Special Issues on results in PISA for different countries. To use databases connected to PISA for research on new topics is therefore seen by the authors as under-emphasized by individual researchers.

Concluding comment

As said in the beginning, this chapter investigates two phenomenon of importance for developing international large-scale assessments – the activity of educational comparisons and the evolution of data usage in education. As showed, educational comparisons are as old as the hills and have for a long time been a fruitful scientific branch within education. But, when statistics as a methodology and a scientific field was developed by the end of the 19th century producing educational data a new way in discussing education, and especially educational results, appeared.

The developments of data lead to a displacement in the style of reasoning: on the one hand in comparative education and on the other in education at large, creating a specific narrative on what education is and what education supposed to achieve. As a result of the interconnectedness between educational comparisons and data usage a particular configuration appeared in the scientific development that enabled researchers with a comparative ambition on student results to address and develop ILSA within a specific comparative discourse. The development of this specific discourse enabled staging comparative international large-scale assessments, which is described in the chapter as practice and policy, also disseminating into regional and national assessments and changing the discourse of society and research in different ways.

This development of ILSA lead to an internal split within comparative education where ILSA to many came to be synonymous with an anti-Humanistic strive away from what was up until then perceived as a core value within comparative education. As such ILSA as a research field came to wander off from the path of comparative education and instead came to develop more in line with social engineering with sayings of efficiency, societal development and economic growth as core values, and as such became more aligned with a policy discourse instead of a traditional science discourse.

SELECTED RESEARCH PROGRAMS

In the following, descriptions of the different ILSA that have been mapped, coded and analyzed will be presented. Part of the reason why these programs have been selected is because they have initiated in quite a lot of debate, both in media and in research. They are also the ones that have got most “hits” when searching for peer-reviewed articles in research journals using the search tool Discovery. As such, it is possible to follow the paths taken by different research-interests, which are easier to visualize with the help of PISA and TIMSS than in the case of CIVED and ICCS.

PISA

The *Programme for International Student Assessment* is a triennial international survey that aims to evaluate education systems worldwide by testing the skills and knowledge of fifteen-year-old students. Up until today students representing more than 70 countries or states have participated in the assessment. In the latest study, PISA 2012, around 510 000 students in 65 countries or states took part in an assessment of reading, mathematics and science. For the next round, PISA 2015, more than 70 countries or states have signed up. PISA is said to be unique because the test is not directly linked to the school curriculum. Instead, the assessment is designed to assess to what extent students at the end of compulsory education can apply their knowledge to real-life situations and be fully-equipped for full participation in society. Additionally, PISA as an ongoing triennial survey also gives the countries an opportunity to compare students’ performance over time and assess the impact of education policy decisions.

Since the year 2000 fifteen-year-old students from randomly selected schools worldwide have taken the test in the three key subjects: reading, mathematics and science, with a special focus on one subject each year of assessment. With a start in the 2012 test some countries also participated in an optional assessment in problem solving and financial literacy. The test, taken by the students, lasts for two hours with a mixture of open-ended and multiple-choice questions that are organized in groups based on real-life situations. Different students take different combinations of the test. Besides the knowledge assessments students and their school principals also answer questionnaires to provide information about the students’ backgrounds, schools and learning experiences, along with questions about the broader school system and learning environment.

While the OECD is primarily concerned with economic policy, education has been increasingly important because it has been reframed as central to national economic competitiveness within a human capital framework, often linked to what is called a “*knowledge economy*.” OECD was founded in 1961 and has since taken on a role as a policy actor – often in collaboration with other international organizations like the UNESCO, the European Union and the World Bank – to promote policy preferences. OECD as a case is interesting because it does not have the legal instruments or the financial muscles to actively promote policy making at the national level within the separate member states. Instead, through ranking, annual reports and national and thematic reviews, the organization actively has chosen to frame policy options for the international agenda of education. As such, PISA has become an integrated part of the international policy agenda in influencing national policy making but also – as is discussed in this report – in how education is interpreted and staged in international educational research.

TIMSS

The latest TIMSS study of 2011 was the fifth assessment in the framework of the IEA *Trends in International Mathematics and Science Study*. TIMSS 2011 assessed student achievement in mathematics and science at fourth and eighth grade, as well as trends over a 16-year period. The previous cycles of TIMSS were conducted in 1995, 1999, 2003 and 2007. All the TIMSS studies have gathered information about the contexts for learning mathematics and science from participating students, their teachers and their school principals as well as data about mathematics and science curricula in each of the participating countries. Combined with these

assessments are questionnaires answered for greater insights into home support and school environments for teaching and learning.

The first TIMSS study was conducted in 1995 where the main data collection occurred in 1994-1995. At this time the acronym TIMSS stood for *Third International Mathematics and Science Study*. The key findings of the study for the primary grades was, besides what can be seen in the league tables, that the gender differences in mathematics achievement were small or even non-existent. Concerning science, boys outperformed girls in about half of the tested countries. In nearly all countries an interest in mathematics and science was reported for both boys and girls. The study could also be used to show that home support and having educational resources strongly influenced achievement. Evident in the study became also the fact that in most countries the most frequently used instructional approaches were working together as a class with the teacher teaching the whole classroom, and students working individually. Textbooks were the major written source that teachers used for deciding how to present the topic. In middle school grades the gender differences were also rather small in mathematics, but pervasive in science. Home factors were also seen as important in explaining differences in achievement in this group of assessed students. These conclusions were also true for students in their final year of secondary school, where there were differences in gender and achievement in science, but not that clear when it comes to mathematics.

The acronym TIMSS reappear in 1999, but this time it stands for *Third International Mathematics and Science Study Repeat*. The target group was students in the eighth grade. The study had two major purposes: to include attitudes in relation to achievement and to compare the achievements in relation to the 1995 study. Concerning attitudes, the study showed that students generally had a positive attitude towards mathematics and science although less so in countries where science was taught in separate subjects at the eighth grade. It was also shown that several countries had increased in average mathematics achievement at the eighth grade, which was also true for some countries in relation to science. Gender differences, still rather small for mathematics in the most countries, was partly in the study explained in terms of boys had a more positive self-concept in the ability to do mathematics and science than the girls, which was reflected in the differences in achievements.

In the year 2003 another TIMSS study was presented: *Trends in International Mathematics and Science Study*. This cycle was said to place more emphasis on questions and tasks that would offer better insight into the analytical, problem-solving and inquiry skills and capabilities of students. The target population was students in grades fourth and eighth. In the 2003 TIMSS study the Asian countries came to outperform many of the other participants. Gender differences were still negligible in many countries for mathematics, but in science at the eighth grade boys had significantly higher achievement than girls in a majority of countries. The home context was also explained as an important factor for describing differences in achievement. It was shown as a clear notion that eighth graders with highly educated parents had higher achievement in mathematics and science. Students speaking the language of the test in their homes were also favored. The study also explicitly discussed that the content also needed to be delivered in the classroom in an effective way.

TIMSS was once more repeated in the year 2007. The population targeted was once more fourth and eighth graders. Again, on top in the test were the Asian countries. Concerning the matter of gender and achievements, some adjustments in the results were noticed. At the fourth grade, gender differences in achievement were negligible in roughly half of the participating countries in both the subjects measured. In the remaining countries the results were rather scattered. Girls had higher achievement in about half and boys had higher achievement in the other half. At the eighth grade girls had on average in the participating countries higher achievement than boys in both mathematics and science. Again, students talking the language in which they took the test at home were favored and so were students with parents who have a higher educational level. Attitudes towards mathematics and science also came out as a strong predictor of the students' achievements. TIMSS 2007 also measured what was called school factors, and one conclusion was that teachers' and principals' positive view of the school climate, high levels of teachers' job satisfaction and high expectations for students' achievement also had students performing better than those with teachers and principals having a lesser focus on these factors.

The latest TIMSS has been conducted is in the year 2011, once again with a target population comprising students in grades four and eight. The study showed that, between 1995 and 2011, many countries significantly improved their overall mathematics and science achievement at the fourth grade. For eighth graders there was

more balance between growth and decline in both mathematics and science. The results in TIMSS 2011 show that the most successful schools in mathematics and science tend to have more affluent students, better working conditions and facilities, and more instructional materials –e.g. such as books, computers, technological support and supplies. A high level of average achievement was also associated with students attending schools that emphasized academic success. This is in the report discussed in terms of rigorous curricular goals, effective teachers, students with a desire to do well and extensive parental support. Another way to express this is that students that attend schools with low discipline or safety problems and who report more frequently on bullying had much lower achievement than their counterparts in safe and orderly schools. Connected to these findings are also the strong positive associations between students' achievement in the fourth grade and their early learning experiences. This result is shown with exemplifications that students having parents that early on engage their children in numeracy activities and enroll them in preprimary education normally do better in achievement tests. This combined with home resources and high expectations by parents explains a lot of the students' achievements. Another important factor in explaining the results is the connection between student attitudes towards mathematics and science and their actual achievements. Notably in this findings are that students attending the fourth grades seem to have a better attitude towards mathematics and science than their counterparts in grade eight.

Civic studies – ICCS and CIVED

The *International Civic and Citizenship Education Study* (ICCS) conducted in the year 2009 was the third IEA study on civic and citizenship education. The first within this field was the 1971 study within the *Six-Subject Study*, and the second was the 1999 *Civic Education Study* (CIVED). The CIVED study was carried out in two phases. In the first phase (1996-1997), researchers conducted qualitative case studies that examined the context and meaning of civic education. The observations from this study were then used to develop instruments for gathering information in the second phase (1999-2000) about students' civic knowledge, attitudes and engagement. The assessment was said to cover the content domains of democracy and citizenship, national identity and social cohesion, and diversity. In the assessment five types of items were measured: students' knowledge of fundamental principles of democracy; students' skills in interpreting political communication; students' concepts of democracy and citizenship; students' attitudes related to their nation, trust in institutions, opportunities for immigrants and the political rights for woman; and finally, students' expectations for future participation in civic-related activities. Additionally, questionnaires were also administered to teachers and school principals, as well as students. The target population included all students enrolled in the grade that contained the most 14-year-old students at the time of testing. In the majority of countries they were in grade eight.

The ICCS study reported on student achievement in a test of knowledge and conceptual understanding, as well as student disposition and attitudes relating to civics and citizenship. For those countries that also participated in the CIVED study overtime, changes in civic content knowledge was measured. Questionnaires for teachers and schools were also gathered for information about the contexts in which students learn about civics and citizenship, including teaching and classroom management practices and school governance and climate. Additionally, a national context survey collected information about the provision of civic and citizenship education in the participating countries. The main data collection was conducted in 2008-2009 among students enrolled in the eighth grade.

A REVIEW OF PISA PUBLICATIONS

In this chapter we will present a review of research based on the OECD *Programme for International Student Assessment*. We will use this review as a “paradigmatic example” and will thus be more detailed here as compared to the other reviews. We will show the way of doing a research review as presented in our design of the review, but now tested in practice. We will follow the steps presented in our design as a paradigm for the reviews in the next chapters. We have already presented the background of the study and the review questions. Thus, we start with how we materialize the scope in terms of definitions and criteria for inclusion of research. This is shown in relation to the filtration process. We present this in detail as well as in the screening process. Important to recognize is that we only included primary research in peer-reviewed journals. This was a way to focus on scientifically recognized research, and thus a way to assess research quality.

Given the screening, we are coding the articles using a specific code sheet for each publication. This coding is organized in tables and is also playing a vital part in the next two steps – the mapping and synthesizing of the research field. We end up with a set of conclusions and reflections on the review process.

Field review and filtration process

Of vital concern in a systematic research review is the inclusion and exclusion of research publications. Below we are presenting how this is done: first by the search engine and then by manual sorting using detailed criteria for what studies to focus on. As is shown, we are interested in research doing international comparisons of school results. Furthermore, we are interested in primary research publications in international scientific journals. Given this, we exclude books, government reports, etc.

Inclusion and exclusion of articles

The articles identified and classified by Discovery are presented and categorized in appendix. Below follows an overview of the filtration process in different steps and the criteria used in each step. We also present a note about which journals the identified articles are published in.

Table 6. Filtration of PISA articles in different steps by Discovery and manual processes.

| The search process of Discovery | |
|--|------------------------|
| Step I | |
| Discovery search: “Programme for International Student Assessment” OR “PISA” AND “education” Limited to AB Abstract and Search mode: Set for Boolean/phrase | |
| 4406 articles | |
| Step II | |
| Limit to ”Peer Reviewed” | |
| 2230 articles included | 2176 articles excluded |
| Step III | |
| Limit the publication date to 1999-2014 | |
| 2225 articles included | 5 articles excluded |
| Step IV | |
| Limit to papers in Academic journals | |
| 2197 articles included | 28 articles excluded |
| Step V | |

| The search process of Discovery | |
|---|---|
| Limit by subject ⁶ | |
| 1811 articles included | 386 articles excluded |
| Step VI | |
| Limit to English | |
| 1662 articles included | 149 articles excluded |
| Step VII | |
| Limit by publication ⁷ | |
| 849 articles included | 813 articles excluded |
| Step VIII | |
| Removal of duplicates by Discovery | |
| 357 articles included | 492 duplicates removed by Discovery |
| Manual process of including and excluding articles (this process is explicit in Appendix I) | |
| Step IX | |
| Manual removal of duplicates based on abstracts | |
| 333 articles included | 24 duplicates removed manually |
| Step X | |
| Manual removal of newsletter based on abstracts ⁸ | |
| 258 articles included | 75 newsletters/grey literature removed manually |
| Step XI | |
| Manual removal of Editorial text based on abstracts | |
| 248 articles included | 10 Editorial removed manually |
| Step XII | |

⁶ Note that the articles are classified more than once, within Discovery. Subjects chosen: programme for international student assessment (N:596), academic achievement (N: 445), education (N:316), foreign countries (N:276), pisa (N:229), secondary education (N:213), organisation for economic co-operation & development (N:198), education & state (N:171), educational change (N:145), literacy (N:123), students (N:122), foreign students (N:117), comparative education (N:116), educational tests & measurements (N:110), education – research (N:107), mathematics -- study & teaching (N:107), science -- study & teaching (N:99), mathematics (N:95), comparative analysis (N:85), educational evaluation (N:85), educational policy (N:84), mathematics achievement (N:83), educational quality (N:80), reading (N:79), elementary secondary education (N:76), scientific literacy (N:74), international education (N:72), social status (N:69), student evaluation (N:64), research (N:60), students -- rating of (N:60), schools (N:56), higher education (N:55), gender differences (N:53), achievement tests (N:50), comparative studies (N:49), learning (N:49), science achievement (N:49), data analysis (N:48), equality (N:47), evaluation (N:46), curricula (courses of study) (N:45), educational equalization (N:45), secondary school students (N:45), cross cultural studies (N:43), reading achievement (N:43), science (N:43), socioeconomic factors (N:43), educational attainment (N:42), teenagers (N:42)

⁷ The following journals have been chosen (note that some of the journals can appear more than once due to slightly similar spellings or presentation. Discovery do not sort these out): Education Journal (N:108), International Journal of Science Education (N:50), Comparative Education (N:40), Educational Research & Evaluation (N:40), OECD Observer (N:38), Education Economics (N:32), European Education (N:31), European Educational Research Journal (N:31), Scandinavian Journal of Educational Research (N:31), Comparative Education Review (N:30), Oxford Review of Education (N:30), Assessment in Education: Principles, Policy & Practice (N:27), Journal of Education Policy (N:24), Economics of Education Review (N:23), European Journal of Education (N:23), Journal of Educational Psychology (N:19), International Journal of Testing (N:15), Oxford Studies in Comparative Education (N:14), Social Indicators Research (N:14), Sociology of Education (N:14), Applied Measurement in Education (N:13), Compare: a Journal of Comparative & International Education (N:13), School Effectiveness & School Improvement (N:13), Educational Studies in Mathematics (N:12), European Journal of Teacher Education (N:12), Educational Research (N:11), British Educational Research Journal (N:10), Science Teacher (N:10), International Review of Education/Internationale Zeitschrift für Erziehungswissenschaft (N:9), Educational & Psychological Measurement (N:8), PLOS One (N:8), Science Education (N:8), Educational Policy (N:7), Educational Psychology (N:7), European Sociological Review (N:7), British Journal of Sociology of Education (N:6), Education & Society (N:6), Canadian Journal of Education/Revue Canadienne de l'Education (N:5), Journal of Adolescent & Adult Literacy (N:5), Psychological Bulletin (N:5), Review of Educational Research (N:5), Cilip Update (N:4), International Review of Education/Internationale Zeitschrift für Erziehungswissenschaft/Revue Internationale de l'Education (N:4), Journal for Research in Mathematics Education (N:4), Journal of Experimental Education (N:4), Journal of Science Education and Technology (N:4), Psychological Reports (N:4), Research in Science & Technological Education (N:4), Sociological Quarterly (N:4), Journal of Education (N:3).

⁸ In the process of reading the abstracts, articles have been removed for being what we call newsletters/grey literature. These are publications mainly with just informative content about PISA. Often they are written by OECD officials for informative purposes.

| Manual process of including and excluding articles (this process is explicit in Appendix I) | | |
|---|------------------------|--|
| Manual removal of misplaced articles based on abstracts ⁹ | | |
| 218 articles included | | 30 articles removed |
| Step XIII | | |
| Manual classification of the abstracts in: | | |
| A: Articles using PISA data for answering research questions | | |
| B: Articles discussing PISA in policy reasons | | |
| C: Articles discussing PISA and methodological implications | | |
| A: 138 | B: 68 | C: 12 |
| Step XIV | | |
| Manual re-classification after reading the available articles | | |
| A: 131 articles included | B: 68 articles removed | C: 19 articles removed |
| Step XV | | |
| Excluded due to availability | | |
| 106 articles available in full text | | 25 articles not available in full text |
| Step XVI | | |
| Manual classification after reading the articles in: | | |
| A1: Comparing/discussing PISA in relation to two or more countries | | |
| A2: Only discussing PISA in relation to one country | | |
| A1: 59 articles included | | A2: 47 articles excluded |

After manual removal of misplaced articles and duplicates, we ended up with 218 articles. Of these 131 were consistent with our search criteria, while the others were excluded. Of the removed were 19 dealing with methodological issues while 68 of them were discussing PISA studies in policy terms. These articles are of interest in other contexts than the specific one we are working with here.

Observations on filtering processes

When doing a systematic research review it was reasonable to focus on primary research publications in refereed journals. As can be noted in the filtering table this first meant that almost 50 percent of articles were excluded. As such, this is a first result of the review: much of current studies are outside research communication by means of scientific journals. It should be noted that the procedures in search for literatures also produced lots of duplicates and errors. Thus, the manual processes for filtering was very much needed in that respect as well. To this is added that several journal articles according to steps X - XIII were not about primary research but editorial texts, methodological issues or analyses of PISA studies in policy-making. Somewhat surprising was the finding that in the last filtering in step XVI, of the 106 articles only 59 were doing international comparative studies and 47 were excluded as one-country studies.

Another important finding is that we excluded several Swedish publications, such as a special issue of *Utbildning och Demokrati* on international comparative studies; these were excluded due to the English language preference. Furthermore, texts such as Yang Hansen et al. (2014) in a report for the *Nordic Ministry Council on Nordic Lights* were not included since they are outside the filtering criteria of peer-reviewed articles. We made additional search by means of the Swedish Search engine Swepub and could identify a few more studies. But the language aspect is very much worth noticing for systematic research reviews of international research.

⁹ These articles contain “Programme for International Student Assessment” OR “PISA” AND “education” in their abstracts, but that is just for discussing these phenomenon in relation to what is actually investigated.

Screening and coding selected publications

In appendix we present classifications of articles and the selection of publications that fit all our criteria. Due to this screening we ended up with 59 articles that had the characteristics we were looking for. These articles were then coded. In the coding process we described every publication in the different aspects according to our coding sheet. In table 7 we describe the publications over different aspects. We will here shortly present and comment on the aspects and what distinctions are used to characterize the studies.

Table 7. Coding selected publications over different aspects. (N:59)

| Aspects | Alternatives | Distribution |
|--|--|--------------|
| What to be inquired? | Student achievement | 49 |
| | School characters | 2 |
| | Education system characters | 8 |
| How to be inquired? | Formal (only PISA) | 46 |
| | Formal plus formal (PISA + other LSA) | 5 |
| | Inside – outside (using PISA for inquiring other social phenomena) and substantial | 8 |
| What research question level? | Individual | 29 |
| | Organisational | 10 |
| | System | 20 |
| | Subject-specific | 0 |
| Type of account | Descriptive | 7 |
| | Explorative – hypothesis-generating | 44 |
| | Hypothesis-testing | 8 |
| What relevance – for whom?* | Professional relevance | 17 |
| | Policy – administrative relevance | 40 |
| | Academic Relevance | 3 |
| | Citizen relevance | 0 |
| | General information | 1 |
| What is the logic from research to practical relevance?* | Directives – how to improve achievements? | 20 |
| | Insight – how to understand achievements? | 39 |
| | Not related | 1 |

* Articles 37A and 137A have been double classified for type of relevance; 37A has also been classified twice concerning the logic from research to practical relevance.

First, we note that in all included cases the studies present themselves as being initiated by the researchers/authors themselves. In accordance with our filtration criteria all studies are based on comparisons between different countries with PISA as its common point for developing comparisons.

Second, we identified the inquiries that were at stake in the publications – what was inquired and how it was inquired. We noted that most of the identified studies were interested in student achievements, though some were centering on school or systems characters. This outcome was to be expected given the focus of our review. As is shown in Table 7 almost all of the analyses are *formal* – they are only based on the specific PISA-material. A few are combining the PISA-data with other large scale data bases – mostly TIMSS – and still others are referring to other data, e.g. indicators on gender inequality across countries or GDP. Another aspect concerns the level of the research question: most inquiries are about individual features, and less often are the publications presenting comparative analyses of how schools are organized, e.g. in terms of teaching or resource allocation, or between national educational systems.

Third, we looked for what kind of accounts concerning knowledge production is presented: Are they having a purely descriptive ambition? Are they explorative as some kind of hypothesis-generating study? Or, is the

research testing explicit hypotheses. These are basic ways of capturing different knowledge interests in research. Somewhat surprisingly, a large majority of the studies are having an explorative approach, as is shown in Table 7.

Fourth, not shown in the table but what can be inferred from above is that there is a variation of knowledge objects in these studies. Often we find an interest in taxonomic classifications of groups in relation to school performances with the ambition to explain achievement gaps over these groups. This aspect will be dealt more thoroughly when we are presenting the mapping of the research.

Fifth, when analyzing issues of relevance, a particular and interesting pattern emerges. We found little research devoted to the academic community. In a way this is an underestimation, since there are research reviews presented in the publications, but when reading conclusions, issues on relevance are mostly addressed within political or administrative frameworks and/or argued being of professional relevance. To this is added that most publications are assumed to provide new insights and that less than half of the studies are presenting directives, for instance on how to improve education achievement by different measures.

Only a few studies are actually referring to causal mechanisms explaining the outcomes of their studies. Instead, most studies present themselves as correlational and as analyzing patterns in order to achieve increased insight, e.g. in how achievement gaps are related to other information. A few other studies are underlining that they, by this kind of research, are unable to make any conclusions about causalities at all. Using the distinction between intensive and extensive research programs (Sayers, 2000) we note that what is done is a search for regularities or significant patterns considering taxonomic groups and their relative achievement.

By means of this coding we have given an overview of the publications and can present a research field whose variance is within sets of principles that give the field a rather homogenous quality in one significant way. It is research with the ambition to identify significant patterns or regularities through correlations that derive from the statistical tools applied, mostly in relation to achievement gaps. Given this, we need to make a more detailed mapping of what kinds of knowledge are searched for and what kinds of understandings are presented.

Observations on the screening and coding processes

This step was time-consuming but turned out to be of big importance in the review. The coding of the publications gave us a first overview of the field of study. Standardized coding sheets developed by us were used to summarize the publications in terms of content as well as the form of the studies which we could turn to in the mapping and synthesizing steps. They were easy to work with and had information that gave quick access to questions that arose when dealing with the specific publications. Given more time for reviewing, it would have been a good thing to return to and revise the coding sheets.

Mapping explanandum and explanans in PISA studies

As presented in the introductory chapter we needed to develop a tool to classify and categorize the variations of research – its ambition and scope as well as kinds of inquiries in order to analyze arguments and styles of reasoning in this field of study. For doing so we use the concepts of explanandum – what is to be explained by the research – and explanans – how, or by means of what, the explanation is to be carried out (cf. von Wright, 1971).

Our analysis is in turn related to the properties in arguments in the research utilizing different kinds of data. The point here is if the results are based on one *inside* kind of dataset – such as PISA – or a combination of datasets – such as PISA and TIMSS – or other kinds of data or facts about the world we live in. Finally we say something about how differences of performances over countries, what international comparisons are ending up in exploring the basis of explanandum, explanans and results.

Again, it has to be underlined that we by explanandum and explanans are not referring to explanations in terms of laws and causalities. In the kind of research that we are reviewing the focus is on correlations between variables, showing patterns and regularities over categories and data. This approach is to our understanding a particular consistency with the ways explanations are dealt with in the current research.

Explanandum

By explanandum we refer to what is to be explained – a fact or a statement such as a mystery. This does not imply that such explanations are due to casual mechanisms, such as specific actions and their results. Rather, what are asked for are analyses of patterns or regularities in order to put forwards conclusions about the strength of these patterns. (cf. Sayer, 2000). In most studies what is to be explained are products of the studies themselves such as patterns of similarities and differences in e.g. school performances.

In PISA studies the notion of “*gap*” is important to the research under review. In the literatures we often find three kinds of gaps: the *opportunity* gap concerns unequal distribution of resources for education or educational opportunities, the *learning* gap refers to the difference between expectations and performances for individual students, and the *achievement* gap refers to inequalities in educational outcomes and benefits among groups – often in terms of socio-economic status, race/ethnicity and gender. We identified that a large number of studies tried to identify and explain achievement gaps as identified mostly by means of individual school performances in various subjects. Common to these “*gaps*” are comparisons between countries, but also between the sexes or other population groups to be acted on, such as ethnicity, social economic status or immigrant status.

To take the achievement gap as one point in developing explanandum is to engage in a particular set of principles about comparison that move across and are internal to the three kinds of gaps. Achievement gaps refer to stable differences between such categories. For instance: what is the achievement gap in reading performances among different countries? And second, what are the achievement gaps within or between education systems, e.g. in terms of SES, gender or ethnicity. Are there different achievement gaps in countries?

What is to be explained are often achievement gaps as measured by performances – in science, mathematics and reading – that are then correlated by pointing to patterns of social and psychological factors, such as student careers, family status, or college attending or drop outs. We identified here three different, but sometimes interrelated, research problematics:

- *Equity Problematics*: Are there biases in education e.g. over gender, social class or ethnicity and why do these occur? This problematic is sometimes connected with the search for imperatives to increase equity in education.
- *Efficiency Problematics*: Are there differences in performances over education systems or schools? Here we find e.g. differences in performances over countries or analyses of measures that are assumed to improve efficiency, such as tracking or teacher performance pay.
- *Direction Problematics*: Here student characteristics are problematic but assumed to be possible to change. How are student expectations or engagement in different subjects and careers to be developed or changed, for instance in relation to science studies and technological occupations?

In these problematics explanandum and explanans are intimately related to each other – they are in a way defining each other. For instance, the equity problematic is defined by a combination of achievement measurements and groupings of individuals in taxonomies.

Explanans

Explanandum and explanans are related to each other in various ways. However, here we present a set of overlapping principles to order the kinds of explanations present. We are not presenting the sequences of explanans-explanandum, since the main point is the basis for explanations. Below is our reasoning about kinds of explanans in relation to data:

- We have *internal* explanans where different variables in the same dataset (that is for a research program) are related to each other – e.g. student homework and reading performance.
- We have *external* explanans, where different data sets are combined, e.g. to capture language comprehension over time by means of different databases. The point is that different measurements are related to each other in order to develop patterns or regularities. Another external explanans are based on the identification of

populational characteristics – e.g. income distribution or indicators of gender equity or kinds migration – that can be related to achievement gaps.

- The logic or principles that relate internal and external explanans are specified, tacitly or explicitly through an abstract model of the school as *system*. *The system as an analytic of thought* is used in order to capture student performances or achievement gaps. The concept of system is not about a social system, for example, but about administrative characteristics of education, e.g. in terms of elements that can be managed through organizational theories about, for example, tracking and differentiation.
- A specific kind of explanans refers to producing greater harmony and efficiency to the system's model through interventions in the *school organization* and school management, resource distribution or teaching.

These different kinds of explanantia can be combined in different ways in order to deal with the explanandum, e.g. taxonomic groups and system organization and so forth, which means that we have to describe and analyze such combinations.

The research results are defined by explanandum in relation to explanans. For instance achievement gaps are (partly) understood as results of differences in school organization. In this kind of studies the patterns or regularities between data are vital – whose significance as understood as measured by the strength of relations in the investigated patterns which can be more or less complicated.

In sum, ILSA research is here regarded as a specific kind of collection, analysis, interpretation and presentation of data. The collected data is information about individuals in different contexts, statements about their individual characteristics and performances on different kinds of tests, combined with information about their education and schooling. The data is analyzed by means of different operations, and the patterns achieved are interpreted in educational terms as is shown below. This is a specific kind of production of statements and of valuating their objectivity within the framework of this style of reasoning. Important to recognize here is that in this research there are limits in going from correlations to causalities and in the relations between indicators and what is indicated. Ways of dealing with these limits are basic in assessing the quality of ILSA research.

An example of mapping practices

We begin with an example in order to clarify how we are dealing with arguments in the review of publication. It is an article by Jerrim & Choi (2014): *The mathematics skills of school children: How does England compare to the high-performing East Asian jurisdictions?* Below we present our summary of the article, based on our coding categories, and then how this in turn is translated into an argument of explanandum and explanans:

How come that England is outperformed by East Asian countries in educational systems? This fact is presented by PISA and TIMSS, but where in time does this gap occur? An inquiry is carried out by comparing cross-national differences in mathematics test scores in TIMSS at age 9/10 years and at age 13/14 years, and I PISA at age 15/16 years. A result of this inquiry – where national differences and student characteristics are considered is that an achievement gap is at place already at age 9/10 between England and the East-Asian countries, but this gap does not increase further on. It is concluded that the achievement gap is not associated with secondary education, but depends on previous education or differences in cultural contexts of the education systems. This is to inform English policy-makers about how to conceive the location of the mathematics achievement gap in order to ensure England's future prosperity and economic success. (Material: PISA 2009 and TIMSS 2003 and 2007. Response ratios are between 78 and 100 percent. Countries: Singapore, Hong Kong, Japan, Taiwan, Lithuania, Russia, England, USA, Italy, Australia, Scotland, Slovenia and Norway.)

This study is then converted by us into the form of explanandum and explanans, leading into a conclusion that is discussed in terms of practical reason:

Explanandum: Differences in school performances between England and East Asia countries according to international comparisons show large achievement gaps. What to do about this?

Explanans: Comparisons of magnitudes of achievement gaps between country test performances over the age of students showing that the gap is already present in the youngest age cohort and does not expand in the years to come.

Conclusion: Reforming secondary education in England does not matter. A tentative conclusion is presented: a cultural shift is needed in England in order to improve school results and economic progress.

Practical reason: Given the conclusion it is not likely that reforming secondary education in England will decrease their achievement gap with the East Asian countries. Earlier interventions are assumed to have a more structuring impact.

This exemplifies our principal way of work: first we give a structured summary of the article based on the coding. After that we convert arguments in the article in terms of explanandum, explanans and conclusion. By means of this we identify different forms of explananda and explanantia with different contents, which then is the basis for the upcoming review syntheses. Our conclusion in terms of practical reason is based on the information given by the inquiry in terms of reduction of uncertainty of an eventual action.

Results of mapping PISA publications

What were the outcomes of using these distinctions in analyzing the selection of PISA research? Our results are based on the mapping of the 59 publications that was identified during the filtering process. We group these results in terms of explanandum and explanans plus the data used for reaching conclusions.

Explananda are referred to achievements on tests – often in combination. Of the 59 publications 23 were classified as referring to results on test concerning competences: 23 in science, 28 in mathematics, and 27 in reading. Furthermore, a large number – 40 publications – were referring to other explananda, such as teacher-student gender matching or self-concept in relation to achievement. Interestingly enough, there are a number of publications dealing with ranking between countries, but these are mostly non-peer-reviewed newsletters and thus excluded here by means of our classification.

Explanantia are referring to different inquiries that are carried out. Of these inquiries 39 were based on different variables in the same data-set (e.g. PISA, 2009) related to each other, such as student homework and mathematics competences while ten more publications were including other data sets, such as TIMSS in defining when in time achievement gaps occur. In 48 or in three of four publications we noted that taxonomic groups are part inquiries, such as gender or SES or combinations of them. This is done to understand achievement gaps or to control for comparisons referring to other explanantia. Furthermore there are inquiries in 30 publications concerning educational systems – how they are constructed, e.g. in terms of when differentiation in different programs occurs or their organization and use of resources, as was put forwards in thirteen publications. To this is added that a number of publications are referring to combinations of inquiries, e.g. taxonomic groups, system characteristics and organizational characteristics, e.g. when relating system inquiries and tracking inquiries in analyzing achievement gaps.

What is then referred to as evidence in these inquiries? First, it goes without saying that all 59 publications refer to PISA – what we here term *inside data*. Of these a smaller set refers to other data, be it TIMSS or PIRLS, or what we here call *outside data*. Second, a third of the publications are also referring to other kinds of information or facts, for instance the proportion of women in the workforce and social inequality in relation to reading performances. This information is often conceived of as facts established in international statistics, but there are also references to theoretically based classifications of welfare state models. And third, a small set of publications are referring to different kinds of evidence in their inquiries in explaining the outcomes of education and schooling.

Observations on the mapping process

The mapping processes were based on the coding sheet but refocused on the explananda and explanantia as could be concluded in the different publications. In the matrices that were developed, different kinds of categories were developed over explananda and explanantia as well as results. What we got were matrices with the different publications as rows and categorizations as columns. By means of this we got a condensed overview that made it possible to group arguments presented in the publications. This was a good thing to achieve, but the price for that was that we had to complement the mapping with re-readings of the articles in the next synthesizing step.

Synthesizing

The synthesis of the ILSA research will be carried out in terms of different explananda. Previously we defined three sets of explananda: Equity Gaps, Efficiency Gaps, and Careers (direction). These were inquired in different ways, referring to different kinds of explanantia in variables and categorizations of subjects and information about educational systems and their organization. Furthermore, we noted that different kinds of competences were in focus – mathematics, science, and reading.

The syntheses will be presented over combinations of explanans and explanandum. Explananda can be formalized in three sets as follows:

- A. Equity Problematic = School Performances + Taxonomic Groups
- B. Efficiency Problematic = School Performances or Knowledge Organization + Educational Measures
- C. Student Direction Problematic = Career Expectations or Interests or Attitudes + Taxonomic Groups

The equity set of explananda is most common – trying to identify biases in education for different taxonomic groups defined by social or cultural backgrounds or gender. The efficiency set is not that common. Here different kinds of educational measures are compared over school performances. Notable is that the direction problematic – what makes persons to become motivated or to pick a specific career or engage in specific educational activities – are occurring in several studies, often with a focus on science and technology.

What are then conceptualized as explanantia to these different explananda? In a broad sense we identified three sets here:

1. Student characteristics
2. Educational measures
3. Contextual circumstances

Students are categorized in a number of ways – as members of different taxonomic groups, as having certain competences or interests – put forwards with the assumption that these characteristics will matter in the analyses.

Educational measures are here based on characteristics about teaching and learning environments as well as distinctions in school organization and educational systems. An important point is that these measures often are combined in multilayered analyses where the composition of school classes and schools are taken into consideration as well as school differentiation and selection mechanisms and tracking within kinds of schools. By systems it is meant the administrative construction of education and how students are moving over time in this system.

In international comparisons a main point concerns the opportunity to compare different national contexts in different ways. We also note that different national characteristics – such as GDP or differences in equity are put forwards in the analyses.

By means of this structuring we are now in the position to empirically identify the coordinates for different inquiries in our mapping and by means of that the opportunities to synthesize these studies.

It should be noted that the explanantia in one way or another is related to student characteristics – as carriers of gender, SES and so on, as well as performances. Not all studies are including other explanantia, but a fairly big share is analyzing the contribution of education measures – such as tracking or amount of homework – and/or context matters – cultures, variation in prosperity – to capturing the explananda problematics.

Table 8. A structure of inquiries in PISA studies. Explananda and explanantia over publication numbers.*

| Explananda | Explanantia over publication numbers | | |
|--------------------------------------|---|--|--|
| | <i>Student characteristics</i> | <i>Education measures</i> | <i>Contextual circumstances</i> |
| <i>Equity problematic</i> | 1003A, 1007A, 1009A, 1015A, 1037A, 1081A, 1103A, 1121A, 1126A, 1132A, 1137A, 1143A, 1149A, 1192A, 1194A, 1197A, 1202A, 1204A, 1206A, 1217A, 1230A, 1238A, 1239A, 1240A, 1253A, 1254A, 1255A, 1260A, 1261A, 1265A, 1266A, 1299A, 1305A, 1325A, 1326A, 1329A, | 1007A, 1009A, 1015A, 1121A, 1126A, 1143A, 1149A, 1192A, 1197A, 1202A, 1238A, 1254A, 1255A, 1261A, 1299A, 1305A | 1037A, 1081A, 1103A, 1121A, 1132A, 1143A, 1192A, 1194A, 1202A, 1206A, 1217A, 1230A, 1238A, 1253A, 1260A, 1265A, 1266A, 1325A, 1326A, 1329A |
| <i>Efficiency problematic</i> | 1017A, 1018A, 1038A, 1078A, 1168A, 1183A, 1205A, 1220A, 1302A, 1342A | 1017A, 1018A, 1168A, 1220A, 1265A, | 1017A, 1038A, 1078A, 1205A, 1302A |
| <i>Student direction problematic</i> | 1039A, 1116A, 1167A, 1185A, 1187A, 1188A, 1189A, 1190A, 1200A, 1212A, 1271A, 1303A, | 1039A, 1167A, 1189A, | 1039A, 1116A, 1185A, 1187A, 1188A, 1189A, 1190A, 1212A, 1271A, 1303A |

* For references to numbers, see appendix 1

In the overview above we find different as well as similar explananda. We also note that the explanantia might be similar or different for similar explananda. Given this, if synthesizing was a trivial process we would expect from a systematic research review that similar explanantia for the similar explananda would provide similar results and conclusions. However, to our understanding research reviews are not trivial – and the same holds true for research as such. We will elaborate this below.

Efficiency problematic research

Considering the efficiency problematic we note that different kinds of performances or competences are in focus for the explanandum. The studies are to a greater or lesser extent focusing on precise research questions. Based on the inquiries, suggestions for policy-makers or for professionals are presented.

In Table 9 we present more condensed information about publications dealing with the efficiency problematic in cross-country comparisons.

We noted a set of publications dealing with the efficiency problematic considering differences in educational measures and how they are associated with similar or different outcomes. Here we focus on mathematic competences and how student achievement is related to different kinds of educational measures.

For instance, considering *homework* Australian students are doing less homework compared to students in Canada (Woods-McConney et al., 2014). What significance then has the time spent on homework? Dettmers et al., (2009) is doing a multilevel analysis based on 40 countries and more than 230 000 students. One point in the argument is the importance of doing such a study which makes it possible to control for confounding variables. They conclude:

At the student level, no clear-cut relationship was established between homework time and achievement across the 40 countries. The results highlight the need to use multilevel analyses and to control for confounding variables in homework research. (Dettmers et al., 2009, p.375)

Furthermore, it can be noted that in twelve of the 40 countries there was a “[...] statistically significant negative relationship between homework time and mathematics achievement” (Dettmers et al., 2009, p.399). In a way Dettmers et al. are doing a study of falsification of the hypothesis of the homework effect. Considering the ILSA as way of work, the authors put forwards a word of caution: “[...] a major problem with nonexperimental large-scale studies examining homework effects is the unclear direction of causal effects” (Dettmers et al., 2009, p.400). So far time for homework seems not to be of much universal significance for a more efficient math education. These conclusions are discussed in relation to previous research and not in relation to policy-making or professional action.

Teachers are part in the education system. For instance, the question on teachers’ pay is put forwards by Woessmann (2011) in a cross-country study on performance pay for teachers in relation to student mathematics achievement. The study involves 190 000 students in 28 countries plus data from an OECD study on teachers’ salaries and pay for outstanding performances (as e.g. in Sweden) in the framework of International Indicators of Education Systems (INES):

The use of teacher salary adjustments for outstanding performance is significantly associated with math, science, and reading achievement across countries. Scores in countries with performance-related pay are about one quarter standard deviations higher. (Woessmann, 2011, p.404)

This conclusion is followed by the warning that this is a study of patterns and that more “[...] work is needed to identify causal relationships underlying the associations reported in this paper” (Woessmann, 2011, p.415).

Table 9. An overview of a selection of research dealing with the efficiency problematic

| Reference | Explanandum | Explanans | Conclusion | Comment |
|-----------|---|--|--|---|
| 1017A | Mathematics performance + nations (England + East Asia) | Comparing 10 + 16 years old students | No increased gap over 10-16 years. i.e. early intervention is better to decrease gap | Combining TIMSS + PISA |
| 1018A | Problem-solving skills and mathematics and science performances | Quality and quantity of arts-related ICT mediated by problem-solving | Quality in ICT matters. Advices of relevance for different groups are presented | Mediating use of problem-solving for mathematics and science performance |
| 1038A | Grafical artifact skills + countries | Unclear argument Cluster analyses | No differences over countries | Exclude |
| 1039A | College graduation rates + countries | Differences in higher education expectations over countries | US low achievers have often unrealistic study expectations | Unrealistic expectations a problem related to myth of meritocracy in the US |
| 1078A | Reading subskills over countries | Comparisons controlling for reading ability, student and school SES | Different differences in reading subskills over countries | Inform reading instruction. |
| 1167A | Mathematic achievement + student attitudes | Comparisons over students attitudes to cooperation | Cooperative attitudes negative for individuals but positive for averages | Social norms and individual preferences important |
| 1168A | Achievement + countries + | Teacher performance pay | Teacher performance pay correlated with | Underlines that this kind of study cannot |

| Reference | Explanandum | Explanans | Conclusion | Comment |
|-----------|---|--|---------------------|--|
| | differences in performance pay | over student SES (ESCS index) | student achievement | consider cause and effect |
| 1183A | Science achievement + student attitudes | Associations of different kinds of interest. | | Contextualized and de-contextualized items |

Another education system ingredient analyzed is the arts-related use of *computers* or ICT resources in relation to students' achievement. Thus, Liem et al. (2014) analyzed the reported quality of ICT use in relation to problem-solving skills and mathematics performance. They conclude:

Structural equation modeling indicated that the quality of arts-related ICT use positively predicted problem-solving skill, whereas the quantity of arts-related ICT use negatively predicted problem-solving skill. Importantly, however, there was an interaction between quality and quantity of ICT use such that the detrimental effect of the frequency of arts-related ICT use was more pronounced among students with low-quality ICT use than for those with high- and moderate quality ICT use. (Liem et al., 2014, p.360)

This finding is regarded by the authors as having important implications for educational practitioners and policymakers.

As can be noted above, the efficiency problematic is covering a range of education measures in relation to different outcomes. Here we used mathematic achievements as a focus and could present studies on teacher pay, timing of system reforms, student use on ICT and so forth. Given this diversity, it would be possible to make a list of advice about different kinds of actions to improve the results of mathematics education.

Conclusions: A first and most important point is that universalistic advices would be highly problematic when translated into practice! The field of study is scattered and interrelations between different research is problematic. Actually a most informative study was the one by Dettmars et al. (2009) on the impact of homework, wherein they state that universal conclusions are not possible. The basis for this is a multilevel analysis where confounding variables are controlled. This is quite in line with the styles of reason in this field of research where more developed analyses are preferred in order to achieve improved and objective knowledge. A second important point refers to the possibilities to present valid statements by means of extensive analyses of patterns. Thus, Woessmann (2011) points out that it is impossible to identify casual directions in the relations studied.

Equity inquiries

A large set of the inquiries is dealing with equity problematics – that is trying to find explanations for this explanandum in terms of characterizations of students and/or education measures and/or context matters. It would take too much space to present a synthesis of all equity gaps. The table below gives an overview of a selection of arguments and conclusions.

Table 10. An overview of research dealing with the equity problematic

| Reference | Explanandum | Explanans | Conclusion | Comment |
|-----------|--|---|---|--|
| 1003A | Engagement in Science + gender + ethnicity | Participation in science activities + SES + formal time | Canada ahead of Australia | Ethnicity here as indigenous and non-indigenous groups |
| 1007A | Achievement gaps + ethnicity | Home + school resources | Larger gap in Australia compared to New Zealand | Indigenous and non-indigenous dealt with differently |

| Reference | Explanandum | Explanans | Conclusion | Comment |
|-----------|---|---|--|---|
| 1009A | Achievement gaps + SES | Early selection + school tracks + school selectivity | Inequalities related to national educational policies | Multi-level analysis over school, country |
| 1081A | Reading skills + countries + student age | Comparing test scores age 4-5 and 15-16 in USA and Canada | US students lag behind already at age 4. Unclear about SES. | NB: poorer SES for US children. Do not blame teachers. |
| 1103A | Mathematics and reading achievement + gender + country | Identifying different gender gaps achievement group and over countries | Presenting complex patterns over performances and age. National gender indicators did not make a difference. | Different gender policy approaches for boys and girls over achievement levels |
| 1121A | AG + Country + gender + school type + SES | Identifies differences over countries | Parental education and school type as unacceptable important sources of inequalities | Latin American Countries ranked in terms of inequality |
| 1126A | Private school attendance + SES + motivation | Private school | | |
| 1132A | Reading, mathematics and science performance + gender + culture | Innate versus cultural explanations National gender differences as cultural arguments | Significant cross-cultural differences in gender gaps. Important to recognize similarities in order to avoid stereotyping. | Focus on the US specific cultural situation. |
| 1137A | Career plans + gender + country | Self-concept + science career plans + culture | Female-male segregation and science self-concept larger in advanced industrial cultures | |
| 1143A | Achievement gap + immigrants + country | SES + cultural capital + family + schools + teachers + welfare state typ | A number of determinants associated with immigrant disadvantage | Societal organization associated with immigrant student achievement |
| 1149A | AG + migrant students + track + school + systems | Student SES affected by school, track or systems | Different GAPS for students in different tracks, school or systems | No system is best for all students. Track level important |
| 1194A | AG over equity and countries | SES (parental education and books and possessions at home) | SES related inequality present in all countries but differ if it refers to parents or to school composition | Advice: combine education policy with other welfare state policies. |
| 1197A | AG over students and system configuration | Student SES (ESCS) and School homogeneity | Different configurations. Delayed selection connected to delayed inequality | NB country or system differences. Identifying configurations |
| 1202A | School segregation | School | Market-oriented systems | A study on |

| Reference | Explanandum | Explanans | Conclusion | Comment |
|-----------|------------------|--|--|---|
| | | differentiation and control over selection.\ | increase school segregation, while public regulated reduce it | segregation and school composition |
| 1206A | AG over students | School choice Selection | Selective schools perform better Choice socially biased. Ability grouping no effect. | NB: the cost of market efficiency vs social bias and disintegration |
| 1299A | AG over SES | SES | SES little correlation but early tracking is important for social inequality | NB: student ability |

Gender matters

Several studies on gender gaps are dealing with differences between male and female students in mathematic competences. In several studies it is noted that the gender gap is not consistent over national or cultural contexts.¹⁰ But, it is also noted that information about indicators of gender inequalities in different national contexts do (Stoet & Geary, 2013) – or do not (Reilly, 2012) – relate to variation in gender gaps in a significant way.

For instance Stoet & Geary (2013) explored gender differences put forwards this argument based on comparisons over 75 countries and 1.5 million students participating in PISA studies:

Across nations, boys scored higher than girls in mathematics, but lower than girls in reading. The sex difference in reading was three times as large as in mathematics. There was considerable variation in the extent of the sex differences between nations. There are countries without a sex difference in mathematics performance, and in some countries girls scored higher than boys. [...] we found no evidence that the sex differences were related to the nations' gender equality indicators.
(Stoet & Geary, 2013, p.1)

The conclusion presented in short above resulted in the following translation into advices to policy-makers on how to deal with gender gaps in education:

The implication is that if policy makers decide that changes in these sex differences are desired, different approaches will be needed to achieve this for reading and mathematics. Interventions that focus on high-achieving girls in mathematics and on low achieving boys in reading are likely to yield the strongest educational benefits.
(Stoet & Geary, 2013, p.1)

Reilly (2012) did a study based on debates on gender differences being innate or derived from cultural and social determinants. Here, international comparisons could serve as important evidence. He put forwards a set of hypothesis to be tested by means of a combination of PISA studies and Gender Equity measures. Data used was the PISA 2009 study including 480 thousand students from 65 countries. He argues that cross-culturally there is stronger evidence for gender similarities compared to the US:

Globally, there is a small gender difference in mathematics literacy favoring males, and a small difference in science literacy favoring girls in non-OECD nations. However, a consistent finding for reading literacy is that girls outperform boys [...]

¹⁰ Stoet, G. & Geary, D. C. (2013); Gamboa & Waltenberg (2012); Sikora & Pokropek (2012); Buccheri, Gurber, & Bruhwiler (2011)

Correlational analyses show that economic prosperity, gender equity, and the dimension of power distance are good predictors of global gender differences in cognitive abilities.

(Reilly, 2012, p.15)

Reilly discuss further social implications of research on gender differences “[...] *that inflated claims of wide gender difference might contribute to increased gender segregation in education and the workforce and that the potential of girls may be overlooked*” (Reilly, 2012, p.14).

Simply stated we have two studies with a similar problematics and inquiries using similar data material, but the conclusions are quite different concerning the size of the measured differences and correlations between school performances and indicators of gender equity. To understand this contradiction in conclusion we need to put it in the light of the following matters. First, what these studies observe are patterns of relations of different strengths and not casual mechanisms. Thus, the directions in these relations have to be left out in the analysis. Second, the two studies have different frameworks. Stoet & Geary (2013) are exploring sex differences in school performances, while Reilly (2012) is testing a hypothesis concerning gender and culture as central concepts. Accordingly, data are interpreted differently. An important difference between the conclusions is that Reilly is reflecting on the risk of reinforcing gender stereotypes by inflating differences in school performances. Stoet & Gary take a different stance by further classifications of students in terms of sex and school performance level. To us this is pinpointing the importance of understanding interaction between research and the researched.

The selected articles are presented in the same journal but in different volumes. Stoet & Gary (2013) noted that Reilly had gotten a different result, but neither discussed this nor their different conclusions. Such findings about research communication are of interest when analyzing a research field but are seemingly not very relevant in terms of practical reason. However, the arguments presented are different and given the fact that Reilly is testing detailed hypothesis and is reflecting on the potential stereotyping effects of research it is reasonable to pay extra attention to that study and neglect the differentiating measures proposed by Stoet & Geary. Such a conclusion is not based on the results of inquiries but on analysis of the arguments presented. Our point is that research synthesizing in this case is based on intellectual evaluations of presented arguments and not about adding results.

Socio-economic matters

Studies on achievement gaps over students’ social and cultural background note that information on social background is strongly related to student performances (Gamboa & Waltenberg, 2012; Mostafa, 2010) and also related to school selection mechanisms such as school markets (Rutkowski, Rutkowski, & Plucker, 2012; Alegre & Ferrer, 2010) and timing for school differentiation (Le Donné, 2014; Lee, 2014; Dronkers, van der Velden, & Dunne, 2012), where early differentiation is increasing the achievement gap. We will present a few studies that are exemplifying this general trend in findings.

For instance, based on analyses of 32 OECD educational systems, Alegre & Ferrer (2010) conclude about market mechanisms in relation to public regulation of school selection:

Results of the regression analyses suggest that more market-oriented school regimes tend to increase schools' social segregation, whilst those characterized as more comprehensive and publicly regulated tend to reduce it. (Alegre & Ferrer, 2010, p.433)

Similar conclusions are presented in several studies. For instance Rutkowski et al. (2012) are analyzing “*cream-skimming*” in relation to privatization:

Instead of traditional cream-skimming, our analysis suggests that a sorting effect of another kind exists – private schools (managed or funded) tend to lure students from better socioeconomic backgrounds, while public schools tend to attract the most engaged students. (Rutkowski et al., 2012, p.375)

By simple head-counting as a means to do a synthesis a reasonable statement is that the privatization and differentiation of students by means of markets is associated with larger social segregation in education.

A related question concerns the timing of differentiation and tracking of a cohort in an educational system. We note here comparisons between national systems that are different in this respect. Thus, Le Donne (2014) carried out a 22 country study on socio-economic equity in education performance in relation to education policy:

We find descriptive evidence that institutional parameters that foster freedom in education, such as an early selection with numerous tracks of study, a great significance of public selective schools, as well as of private schools with fees, jointly amplify socioeconomic inequalities in performances between students essentially by magnifying the effect of schools' social composition on students' competences. (Le Donne, 2014, p.329)

Lee (2014) compared Italy and Austria with late and early timing of school differentiation. This is done in relation to the literature showing that early differentiation is associated with larger socio-economic inequity compared to late differentiation. She analyses educational expectations in relation to this and concludes:

The findings suggest that a later tracking system is associated with higher probabilities of having academic educational expectations; however, this finding should be interpreted with caution as the higher education and vocational education systems are different between the two countries, as well as the valuation of tertiary degrees in the labour market. (Lee, 2014, p.206)

This study confirms what the state of the art is considering early differentiation and higher socio-economic inequity. But there is an important reflection on the results considering the differences in context that might have an impact on the social significance of the study.

A somewhat different study was carried out by Fredriksson, Holzer, McCluskey-Cavin, & Taube (2009), which compared a social democratic welfare model (Sweden) with a liberal model (Switzerland) and their ways of dealing with educational differentiation in terms of selection and tracking. They conclude:

The Swedish system has, with the exception of gender gap in reading, produced a system that seems to have a higher degree of equity than the Swiss system. (Fredriksson et al., 2009, p.54)

Ethnicity matters

Some studies are analyzing achievement gaps over student scategorized as indigenous and non-indigenous. They show differences over national contexts where differences in school organization (Shapira, 2012) and resource distribution (Song, Perry & McConney, 2014) and more over embracing welfare system model (Fredriksson, Holzer, McCluskey-Cavin, & Taube, 2009) are showing significant relations.

As could be noted from these examples are that different explanantia are combined in the inquiries or kept under control in the studies. We also note that there are differences in and demands on more advanced analyses (Dronkers, van der Velden, & Dunne, 2012). Thus, Dronkers et al. (2012) analyze educational inequality for migrant groups in educational systems:

The results highlight the importance of including factors of track level and school composition in the debate surrounding educational inequality of opportunity for students in different education contexts. (Dronkers et al., 2012, p.11)

They continue with a conclusion about how to do such analyses:

The findings clearly indicate that analyses of the effects of educational system characteristics are flawed if the analysis only uses a country level and a student level and ignores the tracks-within-school-level characteristics. (Dronkers et al., 2012, p. 11)

Based on this, a conclusion is presented in relation to education policy:

From a policy perspective, the most important finding is that educational systems are neither uniformly "good" nor uniformly "bad", but they can result in different consequences for different migrant groups. Some migrant groups are better off in comprehensive systems, while others are better off in moderately stratified systems. (Dronkers et al., 2012, p. 11)

A synthesis from this research is first that ethnicity issues are complex with different kinds of classifications and taxonomies at work, though the conclusions are similar to those presented under the heading of socio-economic issues in general. The importance of an analysis that goes beyond a country level and a student level is seemingly important. A more intriguing issue concerns information for research where some countries avoid delivering more detailed information. Thus, Dronkers et al. (2012, p.30) point out that this is not only an obstacle for accurate scientific analysis of migrant students and their educational achievement, it is also socially and politically irresponsible to ignore the importance of country of origin.

To sum up synthesizing matters

We have here showed how we have synthesized research publications dealing with the equity problematics. The conclusions are a bit different considering the different issues:

- *Gender matters*: The gender gaps are different for reading, mathematics and science. And they are different over national contexts as well as in relation to other indicators. It seems to be important to differentiate between gender differences and sex difference when drawing conclusions from research results. The risk for stereotyping gender differences by means of research is to be noted.
- *Socio-economic matters*: there seems to be stable relations between the construction of educational systems and socio-economic segregation as well as performance gaps. Thus, early differentiation is connected with higher socio-economic gaps in performance as well as increased social segregation. Similar conclusions are drawn when analyzing associations between marketization and social segregation. A point to consider is the difference between inside education measurements and the social significance outside the education system.
- *Ethnicity matters*: Taxonomic groups in terms of ethnicity, migration and indigenous students are complicated in many ways. One conclusion is that the distribution of resources on the school level is important for school achievement for indigenous students. Another is that there is a need for elaborated analyses for different migrant groups. An important issue concerns the categorization of students over their home countries.

A side-effect of these studies is the demand for defining taxonomic groups in order to identify problematics and how they should be acted upon. Is it a good practice to categorize students in relation to sex, performance, country of origin, social background and so forth? What is the price of categorization?

Student direction problematics

Under the heading of direction problematics we include studies on student motivation and attitudes to study different subjects or interest in picking certain careers. An overview is presented in Table 11.

Basl (2011) presents worries about lack of interest among university students in natural science related careers. Is it due to family background characteristics or school experience? And how does the Czech Republic compare to other countries? It is concluded:

An almost negligible impact of family on interest formation was found in the Czech Republic, Germany, Finland, and Norway. On the other hand, the impact of school from the point of view of preparing students for future educational and career trajectories was found to be strong in all countries studied due to their influence on the degree of student awareness of science-related career opportunities. (Basl, 2011, p.145)

Underlined in the conclusion is the importance of career guidance in changing student interest in technology and natural science careers.

Table 11. An overview of a selection of research dealing with the direction problematic

| Reference | Explanandum | Explanans | Conclusions | Comments |
|-----------|---|--|---|---|
| 1039A | College dropout graduation rates over countries | Differences in higher education expectations over countries | US low achievers have often unrealistic study expectations | Unrealistic expectations a problem related to myth of meritocracy in US |
| 1116A | Career expectation and gender socialization | Student career expectations given fathers and mothers occupations | Weak intergenerational preference patten that vary over contexts. | Educational implications are discussed |
| 1167A | Mathematic achievement + student attitudes | Comparisons over students attitudes to cooperation | Cooperative attitudes negative for individuals but positive for averages | Social norms and individual preferences important |
| 1185A | Ways that knowledge, affect, and value combine in the structure of students' interest in science might vary in line with historical and cultural traditions | Four countries were chosen to represent contrasting cultural values as defined in analyses | There are different connections between knowledge and value that affect student interest in different context | Notions about curriculum developments are presented. |
| 1187A | Different interests in science issues among students in different countries | Cluster analysis based on PISA interest data | Different clusters were identified where non-European and European countries differ considering student interests | Results consistent with the World Value studies |
| 1188A | Students' interests in science careers over cultural contexts | Defining and comparing different interest in science careers | Country clustering considering interest in different science interests over countries | Cluster studies as a named approach |
| 1189A | Lack of interest in science and technology careers | Comparisons of impact from family and school over countries | Family background not very important. School and career guidance matters most. | Czech republic in focus compared to Finland, Germany and Norway |
| 1303A | Relations between achievement and affect | Separate analyses of 23 countries | Affect and performance are mutually influencing each other. This relation varies over countries. | Conclusion to the research community |

Interest in science studies is analyzed by Sikora & Pokropek (2012) with a focus on gendered intergenerational transfer of career preferences. They present a set of hypotheses considering gendered career socialization. In sum they find:

In many nations, relevant paternal employment enhances sons' interest in science careers regardless of their field. In contrast, maternal employment inspires daughters in fewer countries and this influence tends to be limited to careers in (biology, agriculture and health). (Sikora & Pokropek, 2012, p.234)

This gendered transfer of career preferences is pointing to a need to develop notions of simple reproduction theories as well as the importance to consider contextual differences.

Kjernsli & Lie (2011) present a study analyzing predominating science orientations and interests in science careers, where one background is the problem of recruitment to jobs in this field. They do cluster analyses of student responses to statements about this over a large set of countries. These different clusters are discussed in relation to different cultures and gender stereotypes.

A somewhat different aspect on student interests is presented by Jerrim (2014). He argues that there “[...] is growing concern that many American teenagers hold unrealistic educational plans” (Jerrim, 2014, p.196). Is this something special for the US students with their high drop out from college as a consequence? Given this, he inquires higher education expectations among students with different school achievement over a set of countries:

It shows how expected and actual college graduation rates differ across a number of countries but also that this gap is particularly large in the United States. Additional analysis suggests that this is being driven, at least in part, by the large proportion of low-achieving American children who believe they will go on to obtain a bachelor's degree. (Jerrim, 2014, p.196)

The implications of these findings are discussed in reference to educational policy and contemporary sociological debates as well as concerning the need for comparative studies in this field of study.

The analyses reported here take a position consistent with theory to model the underlying causality of the relationship between affect and achievement as bidirectional. To this end, the present analyses formulate a non-recursive structural equation model which specifies affect and achievement as influences on each other. This model is estimated separately for each of 23 nations, 19 of which are members of the *Organisation for Economic Cooperation and Development* (OECD). All 23 nations participated in the OECD-sponsored *Programme for International Student Assessment* (PISA), a program whose focus is national achievement levels in populations of 15-year-olds. The results of these analyses lend support to the proposition that affect and performance exist in a mutual influence relationship, though the nature of this relationship varies between countries.

Synthesizing from these studies is not easily done. A main point is the worry about student interests, especially in relation to orientations and careers in science. These worries seem to be worldwide, though they differ in direction according to cross-country comparisons. There seems to be divergent explananda considering the impact of family background, e.g. in intergenerational transfer of preferences versus the impact of school measures.

Observations on the synthesizing processes of equity studies

Synthesizing research is at the core in a systematic research reviews, being aggregative or configurative or as in our case argumentative. However, this was shown to be highly problematic, partly due to the research field, partly due to the actual task of synthesizing research for translation into educational practice of different kinds.

Considering the current field of research on international comparisons by means of large scale assessments, it has to be noted that its extensive characteristics have potentials as well as drawbacks.

First, by means of such studies it should be possible to formulate well-grounded statements on education matters in different contexts. However, given the rather small share (a few percent) of studies of all publications (in sum 59 articles on international comparisons in peer reviewed journals of a total of more than 4 000 identified articles dealing with PISA research) such international comparisons are rather infrequent. This is further underlined by the intellectual importance to do strict and developed numerical analyses. It is demanded

that the researchers resist the temptation to make comparisons but go into differences in society as well as in educational measures and their prerequisites according to the style of reason that is developed.

Based on for instance Bunge (1967) presenting the division between operative and substantial analyses such intellectual constraints are further accentuated by the simple fact that there is a predomination of operative analyses (e.g. statistical-mathematical) of special research questions formulated in administrative terms – e.g. achievement gaps or efficiency of different measures. This means that conclusions are often formulated in operative terms. An alternative is research based on substantial analyses where the problematics are conceptually defined and so are the potential contributions of knowledge. To our knowledge there are rather few such studies in the research reviewed here. This point can be stated in several ways, but a thesis to be tested is that the current knowledge production in ILSA is focusing on operative analyses and little of conceptual or substantial analyses.

Second, the ILSA studies are analyzing patterns – or the strength of associations between different variables. To recognize this is an important norm in current ILSA research, which implicates the problem of identifying casual mechanisms, as mentioned in several highly qualified publications as seen in this chapter. Thus, it is highly problematic to define the impact of specific measures – for instance on student achievement.

In sum, there are problems with synthesizing research in the current field for reasons stated above.

Reviewing the PISA research review

In the PISA case we followed the design for a systematic review in the prescribed sets and in accordance with explicit criteria in order to get answers on a set of questions. What conclusions do we make here from the principles and practices that were working when following this design?

First, the *mission* is to review research in order produce insights for the funding and design of the new Swedish institute for educational knowledge (SKOLFORSK). This is something in need to be qualified since this was *not* about collecting specific results, e.g. pros or cons of a specific intervention, or analyzing evidence of, for instance, a certain teaching material. Instead it was a way to analyze a collection of knowledge contributions from a way of researching education and schooling dealing with a variety of study objects and knowledge objects.

Second, the principle to be systematic is here practiced by means of a search machine – the Discovery system that collected thousands of publications. But to rely on the sheer amount of publications or statements is a somewhat doubtful principle – though it looks convincing to state that e.g. “... 43 research publications support the alternative ...,” given this specific kind of educational research, head-counting of research results or findings would not be an advisable practice. Instead, a configurative approach (see Chapter 2; Table 1; and Gough, 2012) with a conceptual focus is more advisable. Stated otherwise, the scientific legitimation of results from a research review needs to be based on the strength of an argument. Being systematic by referring to many publications achieved by search engines is not a good enough principle!

Third, the principles for identifying relevant research were straightforward and easy to present. Research presented in peer reviewed scientific journals was a first principle. A reasonable principle in a systematic research review, we thought, given the amount of journals and peer reviewing as an established practice in assessing research quality. A second principle concerned a scope on empirical comparative studies, that is only research dealing with two or more countries were included. These two criteria for including and excluding research publications seemed to be highly relevant, given our mission and standard practices for evaluating research.

Fourth, our screening and filtering practices also seemed on the spot in relation to our mission and principles for including relevant scientific publications. However, they had implications worthy to consider both in relation to the research field as well as considering the design of the review. Then practicing the criteria turned into an extensive filtration process. We identified in sum 4 406 publication by means of the Discovery search system and ended up in 59 articles – a little more than one percent of the originally identified articles. This extensive filtration was to be expected, given the experience from other research reviews (e.g. Soegard et al., 2013) but is anyhow remarkable and needs to be commented on in more detail.

According to our scope 2 176 publications – or 49 percent – were excluded since they were not published in a peer reviewed journal. An additional 386 articles – or 9 percent – were excluded since they fell outside the subject criteria and 149 publications – or 4 percent – were also excluded due to language issues when running the machine. The last criteria on language is a highly doubtful practice, which we tried to counteract by means of search on SWEPUB, but the outcome was rather meager there, just a little more than ten publications. To our understanding, language criteria for inclusion and exclusion of publications in research reviews are highly important to scrutinize. In the filtering, a limited number of journals (46) were included since they dealt with education issues. In sum, 813 articles were excluded – that is an additional 18 percent of the original 4 406 publications. (These last criteria for exclusion are based on assumed relevance of the publications – which is a bit problematic and should be retested, given our focus on arguments.) After that we have a set of exclusions of more technical – identified duplicates and misplaced articles – in sum 546 texts or 12 percent. The criterion to include only publications that dealt with empirical research made us remove newsletters and editorial texts, together 85 articles.

Given these inclusion/exclusion principles and practices we got a set of 218 articles. If we take away publications that did not deal with empirical analyses of PISA data in comparative analyses and those articles that were unavailable we ended up in the 59 articles that were subject to our review. To our understanding the principle of focusing on peer-reviewed international comparative analyses lead to strict and transparent filtering. But there are two issues that are important to highlight: the principle of scientifically assessed publications excluded a very large share of publications as well as discursive positions. A conclusion from this is that the research field is very much invaded by other aspects of the PISA studies – to their connections to policy discourses and to local inquiries.

A conclusion from this filtration is that PISA research is already contextualized in different ways and that only a small part is devoted to international comparisons of school results. This should actually be considered as a result of the research review. Another conclusion is that there is a problem making limits by means of educational relevance – to develop transparent criteria in focusing on educational problematics. This needs to be further elaborated on in other education research reviews!

Fifth, considering the coding principles and practices, they were based on a set of pre-coded aspects with alternatives on research practices and results and conceptions of research relevance. A set of expected results turned up – a focus on student achievements and research questions on an individual level, but also with an interest in organizational and system levels. However, a set of unexpected outcomes considering the PISA research turned also up. There was strikingly little hypothesis testing in this quantitative research field, and the search was to a large extent addressing professional and political issues and surprisingly few turning to the research community. A bit intriguing was that a very large share of research was only using the PISA data base in analyses and rather seldom made combinations with other databases in their analyses or turned to other kinds of sources of information. To us it meant that PISA research is a kind of research presenting pictures of education by means of “PISA eyes” only, which is important to note in combination with the explorative emphasis in the research carried out. The relevancy was mostly formulated in terms of insights, but there were several studies presenting advices and directives to policy-makers and professionals.

Sixth, the mapping stage was somewhat problematic, given heterogeneity of research objects and inquiries. Analyses of arguments in terms of explananda and explanantia (von Wright, 2004) made this heterogeneity manageable in a way we thought of as stringent enough considering the variation in explananda. Going over to the explanantia there was less variation in the ways the inquiries were carried out but more differences in the levels of analyses (from individual to systems) within the inquiries that were carried. These mapping outcomes made it possible to synthesize research over explananda (equity, efficiency, and student direction problematics) in a way that we conceived of as relevant for this kind of systematic review.

Seventh, when going through the PISA research we noted a kind of homogeneity in the analyses, a set of reasoning considering the population in terms of taxonomies, in reasoning about the validity of certain statements – often related to a more developed analysis in combining different variables and categories. This kind of homogeneity is here regarded as an implication of the style of reasoning (Hacking, 1992b) at work in this field of research. We also noted that some articles pointed to explicit demands on the statements delivered – talking about recognizing the opportunities and limits in the work of analyses of correlations, the problems of

generalizability of research and the problematics in categorizing individuals. However, there was a set of articles ignoring such restrictions in talking about causalities in results and in directives for policy-makers. To our understanding, style of reason is important in research contextualization as well and in much need to be analyzed in research translation and in policy discourses.

A REVIEW OF THE TIMSS STUDY

Field review and filtration process

The systematic research review concerning TIMSS is made in the same manner as the PISA review. Obviously, inclusion and exclusion of research publications is of vital concern in this review as well. We are not describing it on the same detailed level as in the PISA review since the principles and practices are mostly quite similar, though there are some differences in the mapping of research. However, we will often refer to the PISA review.

Below, we present how the review is done, first by the search in Discovery and then by presenting criteria for what studies to include in the review. As we presented before, we are interested in research conducting international comparisons of school results in the same way we were doing in the PISA review. We are also only including research publications in international scientific journals and thus excluding books and governmental reports.

In the same way we described in relation to the PISA study, this systematic way has its drawbacks because we exclude research written in Swedish and we do not collect publications not including education in their abstract text.

Inclusion and exclusion of articles

In Table 12 there is an overview of the filtration process in different steps presented in relation to the criteria used in each step. We started with the identification of potentially relevant articles in Step I and the search words presented there – in all 3391 articles. When introducing the criteria “peer reviewed,” 45 percent of the articles were excluded. An additional 30 percent were removed using the criteria of subject and publications, which is a way to focus on articles being of relevance, given our mission, and still 3 percent more were eliminated due to misplacements, etc.

After manual removals of misplaced articles and duplicates, we ended up with 251 articles. Of these, 182 were consistent with our search criteria while others were excluded. Of the removed, 31 were dealing with methodological issues while 38 of them were discussing TIMSS studies in policy terms. There are some notable differences when comparing the previous PISA study and the review we did on TIMSS studies. The most evident is the fact that fewer of them deal with policy studies in relation to TIMSS than similar articles concerning PISA. On the other hand – methodological issues seem to be more common.

Observations on filtering processes

When conducting our systematic research review concerning TIMSS, we followed the same steps we made concerning the PISA study, with some minor revisions. In the TIMSS study we also found it reasonable to focus on primary research publications in refereed journals. The same structure appeared in the TIMSS study as in the PISA study, where almost 50 percent of the articles were excluded. Therefore, it is also true to conclude with the same statement as in the PISA study that much of current studies and research concerning TIMSS is outside research communication by means of scientific journals. The filtration process we followed also produced lots of duplicates and error. Thus, the manual process for filtering was very much needed in that respect as well. To this is also added that several journal articles, according to Steps X – XIII, were not about primary research but editorial texts, methodological issues or analyses of TIMSS studies in policy making. The same pattern was visible in the TIMSS study: of 140 articles in the last filtering only 53 were doing international comparative studies and 87 were excluded. The bias in that respect is even larger in the TIMSS study than in the PISA study. This means that comparing two or more countries in research is more common in relation to PISA data than it is in TIMSS data.

Another important finding is that we excluded several Swedish publications, such as the already mentioned special issue of “*Utbildning och Demokrati*,” that not only discussed PISA but also TIMSS. This was done due to writing in English as one of our filtering steps. We made an additional search by means of the Swedish

Search engine SwePub and could identify a few more studies. We once more repeat the statement from the PISA study that the language aspect is very much worth noticing for systemic research reviews of international research.

Table 12. Filtration of TIMSS articles in different steps

| The search process of Discovery | |
|---|------------------------|
| Step I | |
| Discovery search: "Trends in International Mathematics and Science Study" OR "Third International Mathematics and Science Study" OR "TIMSS" OR "TIMSS Advanced" AND "education" | |
| Limited to AB Abstract and Search mode: Set for Boolean/phrase | |
| 3391 articles | |
| Step II | |
| Limit to "Peer Reviewed" | |
| 1879 articles included | 1512 articles excluded |
| Step III | |
| Limit the publication date to 1994-2014 | |
| 1877 articles included | 2 articles excluded |
| Step IV | |
| Limit to papers in academic journals | |
| 1829 articles included | 48 articles excluded |
| Step V | |
| Limit by subject ¹¹ | |
| 1553 articles included | 276 articles excluded |
| Step VI | |
| Limit to English | |
| 1507 articles included | 46 articles excluded |
| Step VII | |
| Limit by publication ¹² | |

¹¹ Note that the articles are classified more than once within Discovery. Subjects chosen: Academic achievement (n:465), Mathematics -- study & teaching (n:331), Foreign countries (n:253), Mathematics (n:223), Mathematics achievement (n:219), Education (n:211), Science -- study & teaching (n:208), TIMSS (n:136), Science achievement (n:122), Students (n:115), Educational tests & measurements (n:112), Science (n:103), Grade 8 (n:102), Elementary secondary education (n:99), Teaching methods (n:99), Mathematics education (n:95), Secondary education (n:92), Comparative analysis (n:85), Curricula (courses of study) (n:85), Educational change (n:75), Science education (n:75), Mathematics instruction (n:71), Education -- research (n:67), Comparative education (n:61), Elementary education (n:61), International education (n:59), Learning (n:58), Cross cultural studies (n:57), Educational evaluation (n:57), Programme for international student assessment (n:54), Trends in international mathematics & science study (n:54), Student attitudes (n:53), Cross-cultural studies (n:48), Students -- rating of (n:48), Achievement (n:47), Educational quality (n:47), Instructional systems (n:47), Teaching (n:47), Education & state (n:45), International studies (n:45), Research (n:45), Correlation (n:40), Mathematics tests (n:40), Middle schools (n:39), Mathematical ability (n:38), Mathematics -- study & teaching (elementary) (n:38), School children (n:38), Educational standards (n:37), Achievement tests (n:36), Educational policy (n:36)

¹² The following journals have been chosen (note that some of the journals can appear more than once due to slightly similar spellings or presentation, for. Discovery does not sort these out): Educational research & evaluation (n:90), International journal of instructional media (n:77), Studies in educational evaluation (n:51), School science & mathematics (n:42), Science (n:40), Science teacher (n:40), Educational studies in mathematics (n:36), Assessment in education: principles, policy & practice (n:29), International journal of testing (n:29), International journal of science education (n:26), Mathematics teaching in the middle school (n:20), American educational research journal (n:18), Psychological reports (n:17), Comparative education (n:14), Science education (n:13), Education (n:12), Educational evaluation and policy analysis (n:12), Educational researcher (n:12), Comparative education review (n:11), Oxford review of education (n:11), Economics of education review (n:10), Journal for research in mathematics education (n:10), British educational research journal (n:9), Education economics (n:9), Journal of educational measurement (n:8), Journal of educational psychology (n:8), Compare: a journal of comparative education (n:7), Evaluation review (n:7), International journal of educational development (n:7), Journal of science education and technology (n:7), Journal of marriage & family (n:6), Review of educational research (n:6), American journal of education (n:5), British journal of sociology of education (n:5), Cross-cultural research (n:5), Journal of genetic psychology (n:5), Psychological bulletin (n:5), Social indicators research (n:5), American behavioral scientist (n:4), American journal of sociology (n:4), Annual review of sociology (n:4), Bioscience (n:4),

| The search process of Discovery | | |
|--|--|------------------------|
| 766 articles included | 741 articles excluded | |
| Step VIII | | |
| Removal of duplicates by Discovery | | |
| 317 articles included | 449 articles excluded | |
| Manual process of including and excluding articles (this process is made explicit in Appendix 1) | | |
| Step IX | | |
| Manual removal of duplicates based on abstracts | | |
| 299 articles included | 18 duplicates removed manually | |
| Step X | | |
| Manual removal of newsletter based on abstracts ¹³ | | |
| 265 articles included | 34 newsletter/grey literature removed manually | |
| Step XI | | |
| Manual removal of editorial text based on abstracts | | |
| 259 articles included | 6 Editorial removed manually | |
| Step XII | | |
| Manual removal of misplaced articles based on abstracts ¹⁴ | | |
| 251 articles included | 8 articles removed | |
| Step XIII | | |
| Manual classification of the abstracts in: | | |
| A: Articles using TIMSS data for answering research questions | | |
| B: Articles discussing TIMSS in policy reasons | | |
| C: Articles discussing TIMSS and methodological implications | | |
| A: 182 | B: 38 | C: 31 |
| Step XIV | | |
| Manual re-classification after reading available articles | | |
| A: 182 articles included | B: 38 articles removed | C: 31 articles removed |
| Step XV | | |
| Excluded due to availability | | |
| 140 articles available in full text | 42 articles not available in full text ¹⁵ | |
| Step XVI | | |
| Manual classification after reading the articles in: | | |
| A1: Comparing/discussing TIMSS in relation to two or more countries | | |
| A2: Only discussing PISA in relation to one country | | |
| A1. 53 articles included | A2: 87 articles excluded | |

Child study journal (n:4), Journal of instructional psychology (n:4), Review of policy research (n:4), Journal of learning disabilities (n:3), Lancet (n:3), North American journal of psychology (n:3), Sociological quarterly (n:3), The journal of educational research (n:2)

¹³ In the process of reading the abstracts, articles have been removed for being what we call newsletters/grey literature. These are publications mainly with just an informative content about TIMSS. Often they are written by TIMSS/government officials for informative purposes.

¹⁴ These articles contain "TIMSS" AND "education" in their abstracts, but that is just for discussing these phenomena in relation to what is actually investigated.

¹⁵ Even though the number of not available text might seem rather large, note that many of them should be excluded due to that they only contains data for one country, and as such would be classified as A2. This selection based on abstract level is expressed in Appendix 4.

Screening and coding selected TIMSS publications

After classification of articles and the selection of publications that fit all our criteria, we ended up with 53 articles that had the characteristics we were looking for, and we will now carry out further analysis in different steps using the same procedure as we used in the PISA analysis. We will begin with describing every publication according to our design over different aspects. Before we present the analysis we will comment on the aspects and the distinctions we used for describing the articles.

To start, a similar pattern turns up when compared to the PISA review: the studies present themselves as being initiated by the researchers/authors independently. In accordance with our filtration criteria, all studies are by default based on comparisons between different countries. Later we will present the countries and the comparisons carried out.

Second, we use the concept of explanandum – what to explain – in the same way we used it in the PISA study. Here, we conclude with a similar notion that most publications are dealing with *achievement gaps* in different ways. In comparison to PISA, this is handled more or less in the same way: fronting gender and SES as especially important. In the PISA analysis we noted also that ethnicity was a category in focus of the research. This is not a fact in the same way in the TIMSS publications. Instead, we had to construct alternative ways to discuss and investigate these achievement gaps. In the table below we present an overview of what kinds of research is carried out in different aspects.

The dominating pattern is to capture student achievement, but besides this we found that the studies are analyzing school and education system characteristics. The articles are based on different kinds of data, but in Table 13 we see that nearly all of the analyses are based on *formal* data, meaning that it is exclusively TIMSS being used for answering their research questions. A few of the inquiries are combining the TIMSS data with other national or international large-scale assessments, mainly PISA or other IEA managed assessments. Others are referring to other data, basically the same which can be seen in studies using PISA for the same kind of research questions. The most common data are indicators on gender inequalities or economic development (e.g. GDP). As we already stated in the PISA study, these distinctions are of importance in examining research because they are important in discussing what premises the analyses of patterns carried out and what claims of validity can be further claimed. Another important distinction concerns the level of the research question: are the inquiries about individual features (which they mostly are) or are the publications, as an example, presenting comparative analyses of how schools are organized in terms of teaching or resource allocation between national education systems?

Third, we use the same procedure as we did before, asking what kind of accounts concerning knowledge production are presented. Do the inquiries have a purely descriptive ambition or do they test former research hypothesis or do they generate their own hypothesis? The same pattern can be seen studying TIMSS as in the PISA review: a lot of the articles are explorative, but some differences can be noted as well. The inquiries using TIMSS are thus more concerned with hypothesis-testing than their peers are in the PISA inquiries.

Fourth, as can be inferred from above and from the PISA review, there is a variation of knowledge objects in these studies. Often we have found an interest in taxonomic groups and school performances with the ambition to explain what we call *achievement gaps*. Using the distinction between intensive and extensive research programs (Sayers, 2000), we note that what is done is a search for regularities or significant patterns considering taxonomic groups and their relative achievement. The same can be said to be true for both PISA and TIMSS inquiries.

Fifth, just as in the PISA review we found that TIMSS research also was weakly devoted to the academic community. Similar to the PISA study, this is perhaps an underestimation since there are research reviews present in the articles, but when we read and analyzed the conclusions, issues on relevance are mostly addressed within a political and administrative framework or/and argued being of relevance for professionals. To this is added that nearly all of the publications are assumed to provide new insights, and in relation to the PISA study directives are rather seldom.

The TIMSS studies are just like the PISA studies, in that rather few in actuality refer to casual mechanisms explaining the outcomes of their studies. Instead, they present themselves as correlational and as analyzing

patterns in order to achieve increased insights, and this reasoning maybe be even stronger for the TIMSS studies than in the PISA studies.

Table 13. Describing selected TIMSS publications over different aspects. (N:53)

| Aspects | Alternatives | Distribution |
|---|---|--------------|
| What to be inquired? | Student achievement | 46 |
| | School characters | 2 |
| | Education system characters | 5 |
| How to be inquired? | Formal (only TIMSS) | 43 |
| | Formal plus formal (TIMSS + other LSA) | 5 |
| | Inside – outside (using TIMSS for inquiring other social phenomena) and substantial | 5 |
| What research question level?* | Individual | 20 |
| | Organizational | 14 |
| | System | 16 |
| | Subject-specific | 6 |
| Type of account | Descriptive | 5 |
| | Explorative – hypothesis-generating | 23 |
| | Hypothesis-testing | 25 |
| What relevance – for whom? | Professional relevance | 12 |
| | Policy – administrative relevance | 39 |
| | Academic relevance | 2 |
| | Citizen relevance | 0 |
| | General information | 0 |
| What is the logic from research to practical relevance? | Directives – how to improve achievements? | 8 |
| | Insight – how to understand achievements? | 46 |
| | Not related | 0 |

* Articles 2104A, 2198A and 2217A have been double classified for the type of research question level; article 2171A has also been classified twice concerning the logic from research to practical relevance.

The pattern from the PISA study is once again visible in the fact that the coding has given an overview of a research field with a rather large variance considering what to study. The main ambition of the research is also within the TIMSS studies to identify patterns or regularities through correlations that derive from the statistical tools applied – which is mostly done in relation to *achievement gaps*. This is once again our reason for making a more detailed mapping of what kinds of knowledge is searched for and what kinds of understandings are presented.

Mapping explanandum and explanans in TIMSS studies

We used the same departure as in the PISA study for taking on the TIMSS research by using explanandum and explanans (see previous chapter). As we stated before, explanandum and explanans are not referring to explanations in terms of laws and causalities but a way of organizing arguments and inquiries.

Explanandum and explanans

The notion of *gap* is also important within TIMSS research. We find pretty much the same three gaps that were evident in the PISA studies:

- The *opportunity gap* concerns unequal distribution of resources for education or educational opportunities
- The *learning gap* refers to the difference between expectations and performances for individual students

- The *achievement gap* refers to inequalities in educational outcomes and benefits among groups – often in terms of socio-economic status, gender or subject – or method specific matters.

The same pattern can be found in TIMSS research: a large number of studies tried to identify and explain *achievement gaps* by means of school performance, which for TIMSS means mathematics and science. These gaps were explained by comparisons between countries but also between different taxonomic groups such as gender or socio-economic status. The *achievement gaps* to be explained are often conceptualized and measured as performance that are to be correlated with social or psychological factors. In this we have identified two different, but sometimes interrelated, research problematics that are identical with the ones we found in the PISA study:

- *Equity problematics*: Are there biases in education e.g. over gender, social class or ethnicity and why do these occur? This problematic is sometimes connected with the search for imperatives to increase equity in education.
- *Efficiency problematics*: Are there differences in performances over education systems or schools? Here we find e.g. differences in performances over countries or analyses of measures that are assumed to improve efficiency.

Beside these two problematics identified in both the PISA and TIMSS research, we found a direction problematics in the PISA research, which was trying to investigate how students' expectations and engagement in different subject influenced achievements and how this was related to careers imagined or developed. These studies were more or less absent in the TIMSS study and there was only one study identified that explicitly was trying to explain how unrealistic educational expectations affected careers and applications for higher education (Yu, 2012). Instead we found a considerable amount of studies with a variety in both subjects and in what they were trying to explain (N:16). In this group we could, nevertheless, find some issues grouping them in terms of focus. In this we have first the studies that try to investigate *achievement gaps* in relation to subject-specific matters or specific proficiencies within what is considered to be core knowledge in the subject (e.g. Tatsuoka, Corter & Tatsuoka, 2004; Kelly, 2002). This is investigated from different angles and with different scopes, e.g. how teachers conceptualize mathematics, in a taxonomic diversity based on gender and achievements on item level and differences in different sub-skills. Second, we have studies trying to explain differences in proficiency and issue level based on age of the students participating in TIMSS (Neuschmidt, Barth, & Hastedt, 2008; Wang & Zhu, 2003). In at least one of these studies age is combined with gender differences. Third, we have studies trying to explain differences in student achievement based on societal and cultural differences.¹⁶ These studies are trying to combine these factors with e.g. the amount of poverty in a society, the first and second language of the students, societal attitudes towards education and teachers' competence in terms of differences in teacher education. These studies are very hard to categorize into one single unit, but what is evident is that all the studies are concerned with trying to explain some kind of perceived gap when the data in TIMSS is analyzed. This might also be true to argue for in relation to the equity and efficiency problematics, but these studies are much more focused on a special field for the investigation with common assumptions of equity and efficiency. The last group we identified in the TIMSS articles lacks this cohesion in focus and instead are rather scattered. We are not stating that the research articles themselves lack either common assumptions or cohesion in focus, but due to the minor quantity of articles in each focus we call this the *randomized problematics* when we analyze the articles.

In the same way as in the PISA study explanandum and explanans are intimately related to each other and in some ways defining each other. When conducting the TIMSS study, we note that the set of overlapping

¹⁶ Wiseman, 2012; Hauger, & Sireci, 2008; Kjaernsli & Lie, 2008; Papanastasiou & Papanastasiou, 2004; Vlaardingerbroek & Taylor, 2003

principles are also visible in similar ways as in the PISA study. This means that the same reasoning about kinds of explanans in relation to data is also available to discuss TIMSS studies:

- First we have *internal explanans*, where different variables in the same dataset are related to each other.
- Second, we have *external explanans*, where different data sets are combined. The point is that different measures are related to each other in order to develop patterns or regularities. Just as in the PISA study, another external explanans is based on the identification of population groups that can be related to achievement gaps.
- Third, the logic or principles that relate internal and external explanans are specified, tacitly or explicitly through an abstract model of the school as system. The *system* as an analytic of thought is used in order to capture student performances or achievement gaps. The concept of system is not, just as in the PISA studies, about e.g. a social system but about administrative characteristics of education, e.g. in terms of elements that can be managed through organizational theories.
- Fourth, a specific kind of explanans refers to producing greater *harmony and efficiency* for the system's model through interventions in the school organization and school management, resource distribution or teaching.

Just as in the PISA study these kinds of explanation are combined in different ways in order to deal with the explanandum, e.g. taxonomic groups and system organization and so forth, meaning that we have to describe and analyze such combinations.

The research results in studies dealing with TIMSS are defined by explanandum in relation to explanans. In the same way as in the PISA study, achievement gaps are (partly) understood as results of differences in school organization. The patterns or regularities between data are seen as vital, and the significance is understood as the strength of relations in the investigated patterns, which can be more or less complicated.

Our analysis of research uses TIMSS data as a specific kind of collection, analysis, interpretation and presentation of data. The collected data is information about individuals in different contexts, statements about their individual characteristics and performance on different kinds of tests, combined with information about their education and schooling. The data is analyzed by means of different operations, and the patterns achieved are interpreted in educational terms as is shown in the former chapter concerning the PISA study and below for the TIMSS study. As we stated before – this is a specific kind of production of statements and of valuating their objectivity within the framework of this style of reasoning. We also address limitations, both in the PISA study and the TIMSS study, in going from correlations to causalities and in the relations between indicators and what is indicated. We also state that the ways of dealing with these limits are basic in assessing the quality of ILSA research.

Mapping TIMSS publications

Our results are based on the mapping of the 53 publications that was identified during the filtering process. We have then grouped these results in terms of explanandum and explanans plus the data used for reaching conclusions. In order to be consistent with the PISA review we have only included articles explicitly dealing with performance in mathematics. In total this means 42 of the 53 publications. The remaining nine publications are basically texts only dealing with the TIMSS results in science or other objects of interest not related to mathematics. This will become evident when we present outcomes of our review in this chapter. Noteworthy is that when we discuss grouping into different sets of explanans, the total number of 53 articles is used for discussing and presenting scopes, aims and results. This is especially true for what is presented under the heading of *randomized inquiries*.

In our selection *explananda* refer to achievements on test. Of the 42 publications 26 have been classified as having the core content in relation to the efficiency problematics, 11 in relation to the equity problematics and 5 as having a more randomized purpose directed towards performances in mathematics. The 42 publications in the three categories all have an explanandum, in one way or another, related to student performance. Due to the contents in TIMSS, this is normally only mathematics or science or a combination of these two. Some studies

also have one of these subjects in combination with other subjects (e.g. Helbig, 2012, in this case reading), but they are rare and in these cases TIMSS is combined with other ILSA. The explanandum is a combination of performance compared with performances in one or two other countries. This has to do with our filtering of the articles where articles only discussing the performance in one country are singled out and classified as A2. However, comparisons between countries are evident but can be staged in different ways. Sometimes it is just a comparison between two countries (e.g. Im & Park, 2010), sometimes between groups of countries (e.g. Ce, 2005) or sometimes they compare other nations' results with a specific nation as the comparative part (e.g. Desomone, Smith, Baker, & Ueno, 2005). Other ways of explanandum are when the articles compare performance with other issues such as gender (e.g. Helbig, 2012; Cho, 2012), self-concept (e.g. Mei-Shiu, 2008), socio-economic (SES) status (e.g. Ce, 2005) or school expenditures (e.g. Waldfogel & Zhai, 2008).

Explanantia is, in our study, referring to different inquiries carried out. Compared to the PISA review including other ILSA is much more rare, and of the studies included in our TIMSS selection there is only one study that also includes another ILSA (Helbig, 2012, in this case PIRLS). Taxonomic groups as part of the inquiries are also common, in these cases especially gender and socio economic status, but in comparison to PISA these studies are more concerned with individual and contextual learning factors such as self-perception and learning and cultural differences.

Evidence in the inquiries are normally perceived as the facts emanating from the TIMSS studies, and as such, based on what we call inside data. Some of the studies also refer to what we call outside data, but in comparison to PISA this are rather sparse.

Synthesizing

We will now present the synthesis of TIMSS research with regards to different explananda. Previously we defined three sets of explananda. These were inquired in different ways, referring to different kinds of explanantia in variables and categorizations of subjects and information about educational systems and their organization.

Table 14. A structure of inquiries in TIMSS studies. Explananda and explanantia over publications.

| Explananda | Explanantia over publication numbers | | |
|-------------------------|--|---|--|
| | Student characteristics | Education measures | Contextual circumstances |
| Equity problematics | 2048A, 2061A, 2067A, 2078A, 2110A, 2141A, 2156A, 2198A, | 2034A, 2048A, 2067A, 2110A, 2132A, 2141A, 2156A, 2225A, | 2048A, 2067A, 2078A, 2132A, |
| Efficiency problematics | 2025A, 2030A, 2084A, 2100A, 2103A, 2104A, 2134A, 2157A, 2199A, 2206A, 2208A, 2242A, 2268A, | 2025A, 2030A, 2052A, 2084A, 2103A, 2104A, 2127A, 2140A, 2142A, 2144A, 2145A, 2157A, 2174A, 2184A, 2199A, 2203A, 2206A, 2208A, 2217A, 2235A, 2242A, 2247A, | 2052A, 2084A, 2100A, 2103A, 2127A, 2145A, 2157A, 2206A, 2235A, 2239A, 2242A, |
| Randomized problematics | 2202A, 2020A, 2022A, 2083A, 2086A, 2102A, 2171A, 2175A, 2189A, 2204A, 2227A, 2299A, | 2205A, 2022A, 2029A, 2102A, 2204A, 2299A, | 2202A, 2029A, 2185A, 2192A, |

The synthesis will be presented in combinations of explanans and explanandum. For an overview, see Table 14, where publications are organized over explananda and explanantia.

Explananda can be formalized in three sets (following the PISA study in the two first steps), with a difference in the last due to that the TIMSS research lacking direction problematics, so instead we have grouped studies from a variety of focuses as randomized problematics:

A: Equity Problematics = School Performances + Taxonomic Groups

B: Efficiency Problematics = School Performances + Educational Measures

C: Randomized Problematics = School Performances + Selected characteristics on individual, organizational or system level.

The efficiency set of explananda is most commonly followed by equity and randomized problematics.

What then are conceptualized as explanantia to these different explananda? In broad sense we identify the same three sets here as in the PISA study:

1. Student characteristics
2. Educational measures
3. Contextual circumstances

Students are ambiguous also in the TIMSS research and categorized in a number of ways – as members of different taxonomic groups, as having certain competences or interests – put alongside the assumption that these characteristics will matter in the analyses. Educational measures here are based on characteristics about teaching and learning environments as well as distinctions in school organization and educational systems. An important point is also that these measures often are combined in multilayered analyses where the composition of school classes are taken into consideration with school differentiation and selection mechanisms and tracking within kinds of schools. A contextual circumstance means the administrative construction of education and how students are moving over time in this system. In international comparisons, as already mentioned, a main point concerns the opportunity to compare different national contexts in different ways.

By means of this structuring we are now in the position to empirically identify the coordinates for different inquiries in our mapping and by means of that the opportunities to synthesize these studies. The cases in the TIMSS study have, in this respect, many similarities with the PISA study since explanantia in one way or another is related to student characteristics as well as performances. Besides this, a large amount, but not all, also analyzes the contribution of educational measures to capture the explananda problematics.

In the same way as in the PISA study, we find different as well as similar explananda. We in this study also note that the explanantia might be similar or different for similar explananda.

Efficiency inquiries

We are now working to synthesize research findings where the relation between explanandum and explanans is analyzed while also putting forward conclusions and recommendations. Considering the efficiency problematics we note the same pattern as in the PISA inquiries: different kinds of performances or competences are in focus for the explanandum. Just as in the PISA inquiries, the studies are to a greater or lesser extent focusing on precise research questions mostly directed toward policy makers and professionals. In Table 15 are explanandum and explanans specified in the studies, followed by a summary of the conclusions in the articles and some comments made by us for making the table easier to interpret. For reasons of clarity, it might be expressed that these articles – in relation to the PISA study – are not containing any special section of ethnicity matters. This is because such studies are not evident within TIMSS research.

Table 15. An overview of a section of research dealing with efficiency problematics in TIMSS.

| Reference | Explanandum | Explanans | Conclusion | Comment |
|-----------|---|--|--|------------------------------|
| 2025A | Performance (math & reading) + gender | Teacher/student gender matching | Boys do not benefit, sometimes girls | TIMSS + PIRLS |
| 2030A | Performance (science & math) + gender | Teacher/student gender matching | Little support for gender matching | Other reforms more important |
| 2052A | Performance (math) + nations (US + Korea) | Differences in teachers' instruction | Problem solving the most successful strategy | |
| 2084A | Performance (math + science) + self-concept | Internal/external frame of reference (I/E model) | The I/E model is contrasting in different countries | |
| 2100A | Performance (math + science) + pre-school expenditures | SES + characteristics (school, individual, state) | Small but significant positive effect | |
| 2103A | Performance (math + science) + countries | Self-perception + cultural variances | Positive feedback loop – high self-perception better achievement | |
| 2104A | Performance (math) + countries | Self-belief + classroom instructions | Self-belief and individual work leads to better achievement | |
| 2127A | Performance (math) + countries | Extra Lesson Hours After School (ELHAS) + cultural differences | ELHAS might function for describing differences in achievements | US & Taiwan |
| 2134A | Performance (math) + countries | Self-belief + individual controllable factors (e.g. hard work, studying at home) | Students who attributed success in mathematics to controllable factors (e.g., hard work, studying at home) showed higher test scores whereas students who attributed success in mathematics at school to external factors (e.g., good luck) tended to earn lower mathematics test scores | US & Japan |
| 2140A | Performance (math) + countries (US compared to other nations) | Cultural teaching differences | Need for more conceptual strategies in the US classroom | |
| 2142A | Performance (math) + countries | Cultural teaching differences | The findings show that high quality teaching and learning can take place even in a teacher directed classroom. It is argued that these East Asian classroom practices are deeply rooted in the underlying cultural values of the classroom and the wider | |

| Reference | Explanandum | Explanans | Conclusion | Comment |
|-----------|--|--|--|---|
| | | | society. | |
| 2144A | Performance (math) + countries | Cultural differences in instructions | National patterns of teaching do exist | |
| 2145A | Performance (math) + countries | Cultural differences | Differences in knowing different content | US, Singapore, Israel |
| 2157A | Performance (math + science) + westernized/Asian countries | SES + self-perception + school and classroom environment + out of school activities | Showing differences between the US and Asian counterparts which effects achievement | US and 5 Asian |
| 2174A | Performance (math) + countries | Teacher instructions + private tutoring | Teachers credentials is not having an effect of cognitive skill development | Japan, Korea, US & the Netherlands |
| 2183A | Performance (math + science) + resource usage | School policies (both quantity and quality) | School policies seem to be more important than resources in explaining achievement | |
| 2199A | Performance (math + science) + self-perception | High/low academic standards | High performing countries have high academic standards, low performing – low academic standards | |
| 2203A | Performance (math) + curriculum alignment | Cultural differences in intended, potentially implemented and implemented curriculum | The many differences demonstrated make it clear that there is more than one way to do 8th-grade mathematics | |
| 2206A | Performance (math) + Asian prerequisites for achievement | Cultural prerequisites | Asian students (Hong Kong, Japan, Korea, Singapore) perform better due to cultural values | |
| 2208A | Performance (math) + attitudinal + instructional variables | SES + math attitudes + peer influence + computer use + small group work + extra math lessons | 4th-grade students who liked math, who thought it was important to do well in math for their friends, who did not need private instruction, and who had not been taught using computers or small group procedures, tended to be the better students | Cyprus, Hong Kong, US |
| 2217A | Performance (math) + class size | Characteristics of school, teacher, classroom, system | The TIMSS data reveal that eighth-grade math class sizes vary according to the system of educational governance. Not only are math classes larger in size in centralized school systems than they are in decentralized systems, they are also more homogenous in the centralized system than in the decentralized systems. | Australia, Canada, France, Germany, Hong Kong, Korea, Iceland, Singapore, and the United States |

| Reference | Explanandum | Explanans | Conclusion | Comment |
|-----------|---|--|---|---------------------|
| | | | Interestingly, math achievement is higher in the former than in the latter systems. | |
| 2235A | Performance (math) + gender | Differences in sociocultural practices | Gender mathematical abilities are posited to emerge only with prolonged exposure to certain specialized sociocultural practices | |
| 2239A | Performance (math + science) + countries | Part time employment | For boys, and to a lesser extent for girls, there are negative effects on math and science achievement that are associated with adolescent employment | |
| 2242A | Performance (math + science) + self-perception | Cultural differences | Countries with higher student self-evaluations usually performed poorly on the TIMSS tests and vice versa | |
| 2247A | Performance (math) + countries | Calculator use | Statistically significant negative relationship between students' frequency of calculator use and student performance in Japan; no statistically significant relationship was detected for the US and Portuguese samples. | Japan, US, Portugal |
| 2268A | Performance (math) + factors behind achievement | SES + attitudes | SES and attitudes was after modelling found to be the most important factors | |

What becomes evident in the table is the fact that most of the inquiries, in some way or another, focus on student performances. In this it becomes visible that the performances are stated in terms of what TIMSS is measuring; in many cases this interpretation is taken for granted. In the articles categorized as A1 this is most common. If we had chosen to analyze articles classified as A2, B or C another picture could emerge. In this question of what performances actually are, much more is investigated and discussed in terms of what knowledge can be highlighted in terms of more than just performances on TIMSS tests.

When explanans is discussed a more scattered picture emerges than what is visible in the analyses of PISA. Even though a lot of the explanans is the same, it can be stated that TIMSS' inquiries are more outspoken on cultural differences and schooling practices than their counterparts in PISA when investigating the problematics of efficiency in relation to performance. In the PISA inquiries gender and SES are the most investigated explanans; this is also common among the TIMSS inquiries, but these are, to a larger extent, combined with the cultural and schooling practice differences. Below we will present some of the most used explanantia for explaining differences in the performances within TIMSS. This is presented in three different sections.

Gender matters

Studies within the efficiency problematic explicitly uplift gender issues. This is done from more or less equal positions – trying to inquire how teacher/student gender matching affects performances. The article presented in *British Journal of Sociology of Education*, written by Helbig (2012), approaches the field by discussing around the fact: “*The prevalence of woman in the teaching profession has been claimed by various scholars to be responsible for the low performance among boys*” (p.661). In the other article published in *Economics of*

Education Review, by Cho (2012), the same matter is approached by stating, “*While some educators argue that teacher – student gender matching improves student performance, there is little evidence to support this hypothesis*” (p.54). Concerning the results, the studies are also very close in their conclusions. Helbig (2012) finds that boys do not benefit from being taught by male teachers, neither in mathematics nor in reading. The results show however that girls in some countries seem to profit from being taught by female teachers. In the study by Cho (2012), this conclusion is repeated with the difference that the small positive effects for girls are not mentioned.

Socio-economic matters

Studies within the efficiency problematics, discussing SES as an explanans for mathematics performance is much more common than studies discussing gender. Similarities are found in the fact that SES is evident for explanans in explaining the differences in performance. These similarities are however combined with other sets of explanans, such as self-perception/attitudes and characteristics other than SES of individuals, schools and states. In one study presented in *Educational Research and Evaluation*, by Shen (2005), America is compared with five Asian countries that do better in TIMSS 1999. The aim of the study is to diagnose American schools for better efficiency with the help of TIMSS data. The differences in America are summarized in the following way:

- A relatively shorter school year;
- a higher student body mobility;
- more absenteeism and frequent class interruptions;
- students spend more time watching TV, playing sports, and working on paid jobs;
- a higher percentage of students are from single-parent families;
- on average, parents have a relatively higher educational background;
- a higher percentage of students have computers at home, a lower percentage have their own desks;
- a higher percentage of American eighth graders think they are doing well in mathematics and science and perceive the two subjects as easy. (Shen, 2005, p.195)

The significant differences are then said to provide information about the weaknesses of the American school system, and as a conclusion the author says that “[...] *the most important implication for American middle schools is the need to restore the central position of schooling for American adolescents*” (Shen, 2005, p.195).

In another study presented in the same journal as above, Papanastasiou (2002) is trying to find factors other than SES that differentiate mathematics students in Cyprus, Hong Kong and the USA. In this the author found that individual characteristics like students who liked math, who thought it was important to do well in math for their friends, who did not need private instructions, and who had not been taught using computers or small group procedures tended to be the students with better performance.

These studies by Shen (2005) and Papanastasiou (2002) can be illuminated as two counterparts in the articles dealing with SES within efficiency problematics. The first study is rather close to the data presented in TIMSS while the other is more elaborate with the findings and is trying to develop different ways of using the data to be able to go beyond the first statements that SES is the most important explanans of mathematics performance. The SES matters can also be compared with what we found in the PISA study, where SES is more frequently used as a dominant indicator for predicting performances. In the TIMSS studies SES is handled more elaborately and is often used as a first stepping stone to probe explanations of differences in performance. One

explanation of this is that TIMSS research seems to be more occupied with trying to be near the students, in the classrooms and studying the learning processes. This is what we describe in the next section.

“Inside the black box” matters

Of the studies categorized in handling the efficiency problematic we have found that most articles in some way or another try to go beyond SES or gender as the dominant explanans. Instead these studies are “entering” the classrooms to study processes or methods of learning, teaching instructions and organizational or cultural differences for explaining differences in performance. These articles can, on a rough level, be categorized as using the explanans of self-perception/self-belief, teaching differences, methods in use and school policies.

There are some studies discussing achievements in relation to self-belief and self-perception. When going through the available studies in relation to ILSA (PISA, TIMSS and CIVED/ICCS) the interest for explaining performance based on varieties in self-belief or self-perception is rather large. Most of the studies conclude that if students have confidence in themselves their results will be better. When we analyzed the studies we found that studies discussing these issues on the surface seemed to be equivalent in scope, aims and results, but when going deeper into the methodological parts of the studies we found that what is classified or aggregated as self-belief and self-perception, with the help of TIMSS data, differed in the various studies. In fact, the aggregated self-belief and the self-perception were combined with student characteristics, educational measures and contextual circumstances. This means that studies are investigating characteristics among students that can capture performances – such as willingness to do homework, interest in a subject or career plans knowing that the specific knowledge is necessary. Educational measures are focused on teachers’ abilities to promote interest among the students and different methods in use for increasing their interest. And finally, contextual circumstances – such as cultural differences in promoting education – are investigated in relation to promoting students’ willingness and interest in studying.

Studies using TIMSS data for trying to explain differences are also commonly interested in cultural differences in teaching and the methods in use. As mentioned above, these can be connected to discussions about self-belief and self-perception, but they can also be presented as separated from these discussions. One discussion is made in Wu & Ercikan (2006) where they investigate differences in differential item functioning (DIF), meaning that they try to find factors behind why students have differing results in relation to various items in the TIMSS test. In this they investigate different sources of DIF, and the one especially uplifted in the study is the importance of the context of extra lessons after school and how this affects students’ possibilities to solve certain items. Other studies, such as Im & Jin Park (2010), study what kind of methods leads to best performances. In their study they compare US and Korean students and conclude that the most helpful strategy was encouraging students’ independent problem solving. Another study is investigating achievement and calculator use. This study is a comparison between USA, Japan and Portugal (Tarr, Uekawa, Mittag, & Lennex, 2000), and while the study shows that no significant relationship was detected for the US and Portuguese sample, in the Japanese sample a negative relationship was found. Several explanations are presented in the study but at least one is dealing with cultural differences in how the teaching with calculators is perceived by both students and teachers.

The last group of articles that we present within this section is studies that are trying to explain differences because of differences in school and national policies. This group of articles is widely spread in scope and interest, but what combines them is an interest in achievements dependent upon curriculum. As such their aim is to analyze different national curricula and policies and from this derive some conclusions that make it possible to explain differences in student performances. The group of articles is rather small; however, it is interesting to note that at least one (Cogan & Schmidt, 2002) uses the distinctions developed within IEA measurements between intended, implemented and perceived curriculum.

To sum up:

We have here showed how we have synthesized research publications dealing with efficiency problematics. The conclusions are a bit different considering the different issues:

- *Gender matters*: Within TIMSS research comparing two or more nations, the explanans of gender is in our selection used for discussing teacher/student gender matching.
- *Socio-economic matters*: TIMSS research is more open than PISA research in searching for alternative explanantia connected to SES. Even though SES is considered to be most important for explanans of the performances in mathematics, a lot of the studies also find other indications of importance in discussing gaps within the efficiency problematics.
- *“Inside the black box” matters*: Compared to PISA research there is more interest in TIMSS research on efficiency for issues closer to teaching and learning practices.

Equity inquiries

As noted above we could show some similarities and differences between the PISA and TIMSS inquiries dealing with the efficiency problematics. The same can be said to be true concerning the equity problematics. In Table 16 we have gathered the TIMSS inquiries investigating especially mathematics performance in terms of explanandum and what kind of explanans that is in use.

Several of the articles are dealing with equity problematics in terms of mathematics performance – that is trying to find explanations for this explanandum in terms of characteristics of students and/or education measures and/or context matters. Instead of giving a full synthesis of all the equity gaps visible in the research, we are forced to summarize these in some way or another. Table 16 is made for an overview of a selection of arguments and conclusions within the articles explicitly discussing mathematics achievements. We are following the same pattern as we used in the PISA study. It shall, for reasons of clarity, be said that compared to the PISA study there are no studies presented on ethnicity matters. The reason for this is that no such studies were found in our selection dealing with the equity problematics.

Table 16. An overview of TIMSS research dealing with the equity problematics.

| Reference | Explanandum | Explanans | Conclusion | Comment |
|-----------|---|---|---|-----------------------------------|
| 2034A | Performance (math) + countries | Differences in teachers' preparation for teaching, professional development, perception of school environment | Curriculum, pedagogy and evaluation do effect performance | Saudi Arabia + Taiwan |
| 2048A | Performance (science + math) + family structure | SES + characteristics (state, school, parents, individual, cultural) + family composition | No single policy appear to have large effect | Us compared with 13 other nations |
| 2061A | Performance (math) + countries | Gender + attitudes | Gender equity in school enrollment, women's share of research jobs, and women's parliamentary representation were the most powerful predictors of cross-national variability in gender gaps in math | |
| 2067A | Performance (math + science) + countries | School + SES | SES is more important than school in explaining variations in performance | |
| 2078A | Performance (math + science) + countries | Gender + state characteristics | Differences among high achievers are related to gender inequality in the labor market and in the overall status of men and women | |

| Reference | Explanandum | Explanans | Conclusion | Comment |
|-----------|--|--|---|---------------------------|
| 2110A | Performance (math) + countries | Teacher quality + SES | The data show that many countries that have achieved an overall high quality teaching workforce not necessarily offer equal access to qualified teachers across students of various socioeconomic statuses | |
| 2132A | Performance (math + science + reading) + countries | Mass media and computers | Access to mass media and computers may increase achievement | |
| 2141A | Performance (math + science) + countries | SES + resources and institutional settings | The results show substantial effects of student background on educational performance and a much lower impact of resources and the institutional setting | 7 East European countries |
| 2156A | Performance (math) + countries | SES + instructional environment | The variability in performance is partitioned within and between learning environments, and in the way that achievement is related to background characteristics | |
| 2198A | Performance (math) + economic development | SES (especially in terms of single/double parent families) | SES important in explaining differences in achievement, single parent important as an indicator of marginalization | |
| 2225B | Performance (math) + shadow education | Educational resources and funding to education | Results show that institutional factors of education, including limited access and lower levels of funding, drive the use of shadow education, instead of high-stakes testing and national achievement incentives | |

Gender matters

Studies dealing with gender matters in terms of equity are less than what was found in the PISA review. There are only two studies (Else-Quest, Hyde, & Linn, 2010; Penner, 2008) that are explicitly stating that this is the prime focus in their research. However, there are more studies that in one way or another touch the issues without being explicit on the matter. This is e.g. done in one of the articles from the research question of “*What factors might contribute to gender differences or similarities in math achievement, attitudes, and affect?*” (Else-Quest, Hyde, & Linn, 2010, p.106). The question is then investigated in relation to the gender stratification hypothesis, meaning that gender differences are closely related to cultural variations in opportunity structures for girls and woman. The data in use in the article is interesting for our systematic review since they use both PISA and TIMSS; consequently, they are visible in both of our datasets. As such, the article is interesting because it states something exclusively about the gender gaps with both PISA and TIMSS but also in comparisons between the two. When presenting the results from comparing results in TIMSS and PISA it is said:

Eighteen nations participated in both TIMSS and PISA; we compared gender differences on the math composites from the two assessments. Gender differences on the two assessments were not significantly correlated in magnitude ($r = .27, p > .05$), though lack of statistical significance is probably due, in part, to the small sample of nations participating in both assessments. Paired-

samples t tests demonstrated that the differences between the TIMSS-Math and PISA-Math assessments are not substantial (mean difference = 0.05, SD = 0.11, $t = 1.92$, $p > .05$). [...]. As demonstrated, their distributions overlap significantly. Items from the TIMSS scale of Self-Confidence in Math and the PISA scale of Self-Concept are very similar; both include items reflecting students' perceptions of their ability to do well in math and to learn math quickly. Thus, as with the math composites, we conducted analyses to compare the effect sizes in nations that participated in both assessments. Gender differences on the two assessments were highly correlated ($r = .77$, $p < .001$). Yet, paired-samples t tests demonstrated that gender differences on the two assessments differed significantly within nations (mean difference = 0.08, SD = 0.09, $t = -3.81$, $p < .001$). (Else-Quest, Hyde, & Linn, 2010, p.1017)

What becomes visible in the referenced part above is the high degree of statistical calculation for saying something about gender gaps in PISA and TIMSS and how this might be interpreted in relation to the gender stratification hypothesis. The other article is also trying to explain gender gaps from a societal level of explanations. The article is introduced by saying that "*The underrepresentation of woman in mathematics and science has been the subject of much academic debate*" (Penner, 2008, p.138). From this stepping stone the article continues:

Genetic and other biological explanations have reemerged in recent scholarship on the underrepresentation of women in mathematics and the sciences. This study engages this debate by using international data—including math achievement scores from the Third International Mathematics and Sciences Study and country-level data from the World Bank, the United Nations, the International Labour Organization, the World Values Survey, and the International Social Survey Programme - to demonstrate the importance of social factors and to estimate an upper bound for the impact of genetic factors. (Penner, 2008, p.138)

The article then concludes:

While the international variation documented in this study highlights the continued relevance of social factors, given the multitude of biological processes required to take a mathematics test (e.g., reading the problem, thinking of the solution, recording the answer), it is hard to imagine that differences are not at some level biologically mediated. Thus, as future research increases our understanding of the biological processes involved in mathematics achievement, it is important to remember the ways in which these processes are shaped by social factors. To that end, this study suggests that future research would do well to examine the influence of macrosocial contexts; in particular, research examining how macrolevel factors interact with microlevel factors (e.g., parent, teacher, and peer effects) to form gender inequality across the distribution would be informative. In addition, the international perspective is potentially helpful in increasing our understanding of how biological and social factors interact more broadly. For example, studies of gender differences in cognitive abilities might examine how national differences in factors like diet interact with considerations such as hormones and iron deficiency to impact gender differences. Or in the case of genetics, there is little international research on the heritability of cognitive abilities [...], and work that systematically examines how heritability estimates vary internationally would be of great interest. Similarly, while its complexity should not be underestimated, research attempting to identify combinations of genes contributing to intelligence [...] and how they interact with environmental factors could also benefit from a cross-national perspective. (Penner, 2008, pp.164-165)

In this case we have two similar studies trying to explain gender gaps in achievements using similar data, but the conclusions are quite different concerning focuses on what is to be explained and discussed. While Else-Quest, Hyde, & Linn (2010) is rather faithful to the data presented within PISA and TIMSS and basically is just doing re-analyses of the material, Penner (2008) is trying to discuss gender differences with the help of a multitude of indexes for investigating different opportunities dependent on gender. In this Penner is close to a biological or genetic discussion even though he is not exactly within this branch of research. The width of these two studies can serve as an exemplification of the variety of studies that TIMSS data leads to. A note can also be made that neither of these articles are published in explicitly educational journals. The first article is published in *Psychological Bulletin* and the second in *American Journal of Sociology*.

Socio-economic matters

Studies on achievement gaps over students' social and cultural backgrounds are much more evident within the equity problematics than gender; no less than six articles out of eleven discuss these issues.¹⁷ This is maybe not that surprisingly considering that SES might be discussed as a prerequisite for a lot of analyzes of equity. More surprisingly was the low amount of articles on gender. The articles with a focus on SES also related to other issues for discussing equity, such as characteristics of individuals, schools and states (Hueveline, Hongxing & Timberlake, 2010; Chudgar & Luschei, 2009; Akiba, LeTendre, & Scribner, 2007; Ammermüller, Heijke, & Wössmann, 2005), instructional environment (O'Dwyer, 2005) and family composition (Hueveline, Hongxing, & Timberlake, 2010; Schiller, Khmelkov, & Wang, 2002). Below we will present some of the studies that exemplify the general trend in our findings.

For instance, Hueveline, Yang, & Timberlake (2010) conclude in an overall statement concerning family composition, as an indicator of SES, and student performances that:

Results from 2-level hierarchical linear models demonstrated that international differences in the educational gap were associated with several indicators of national policy and demographic contexts. No single policy appeared to have a large effect, but several policy combinations were associated with substantially reduced educational gaps between children from different family structures. (Hueveline, Yang, & Timberlake, 2010, p.1362)

The other article concerned with family composition (Schiller, Khmelkov, & Wang, 2002) and student performance concludes with a similar idea:

[...] our results suggest that the influences on mathematics achievement of some aspects of adolescents' social background are more sensitive than others to national economic conditions. Taking into account individual differences in academic ability and orientation toward schooling, both parents' education and living in traditional families were positively related to higher mathematics test scores in these 34 nations participating in TIMSS. However, we found no discernable pattern across nations in the effect of parents' education on middle-school students' mathematics achievement related to national levels of economic development. In contrast, both our OLS regression analysis and HLM results suggest that the effect of family structure was significantly stronger in nations with higher GDPs. (Schiller, Khmelkov, & Wang 2002, p.740)

As a conclusion it can therefore be stated that family composition as such is not enough to explain differences in performance. Instead, both articles also involve other individual and contextual factors for explaining the results in TIMSS. Both these articles are presented in the same journal – *Journal of Marriage and Family*.

At least two articles uplift instructional environments for explaining performances (O'Dwyer, 2005; Dodeen, Abdelfattah, Shumrani, & Hilal, 2012), but only one is explicitly relating the issue to SES. In this article (O'Dwyer, 2005) the author concludes:

This research highlights the differences among countries both within and across TIMSS administrations, not only in terms of average mathematics achievement and total performance variation, but in the way that the variability in performance is partitioned within and between learning environments, and in the way that achievement is related to background characteristics [...] This finding is similar to previous examinations of the effects of ability tracking on achievement which found that tracking is not necessarily associated with higher performance. These results suggest that tracking, or the lack of it, is not a panacea for improving student performance. Despite the lack of any strong observed association with achievement in these and other data, previous research has found that tracking tends to reduce the positive peer effects that low-ability students might gain from being exposed to more able students [...] tends to benefit higher ability students at the expense of less able ones [...], and to create unequal opportunities for academic success which have the potential to lead to future wage inequalities. (O'Dwyer, 2005, p.174)

¹⁷ Hueveline, Hongxing, & Timberlake, 2010; Chudgar & Luschei, 2009; Akiba, LeTendre, & Scribner, 2007; Ammermüller, Heijke, & Wössmann, 2005; O'Dwyer, 2005; Schiller, Khmelkov, & Wang, 2002.

In this the article concludes that achievements are related to SES, and that tracking is not a way to deal with these differences, for tracking instead creates unequal opportunities that might lead to future wage inequalities. The article is published in a journal well established within education – *Educational Research and Evaluation*.

The last group of articles observed in how to explain and discuss achievement gaps in terms of equality includes articles connecting differences in SES to individual, school and state characteristics. This is the largest group of articles all striving for explaining complementary factors to SES affecting performances.

The questions asked are somewhat different. In an article published in *American Educational Research Journal*, by Chudgar & Luschei (2009), they open with a statement, “*The international and comparative education literature is not in agreement over the role that school resources play in learning*” (Chudgar & Luschei, 2009, p.626). With this as a stepping stone, the authors reexamine this debate by doing a research review across 25 countries participating in TIMSS 2003 and finding that:

(a) In most cases, family background is more important than schools in understanding variations in student performance; (b) schools are nonetheless a significant source of variation in student performance, especially in poor and unequal countries; (c) in some cases, schools may bridge the achievement gap between high and low socioeconomic status children. However, schools' ability to do so is not systematically related to a country's economic or inequality status. (Chudgar & Luschei, 2009 p.626)

This can be compared with an article in the same journal, by Akiba, LeTendre, & Scribner (2007), that is more concerned with connecting SES and teacher quality for explaining performances. In this article the dilemma of explaining unequal performances is set in the agenda of “*Unequal access to high-quality teachers is, of course, part of a larger question of differential access to education based on socioeconomic, racial, or other background attributes*” (Akiba, LeTendre, & Scribner, 2007, p.370). With this, the discussion concludes that:

Cross-national analyses revealed that the countries with better teacher quality produced higher mathematics achievement. However, larger opportunity gaps in access to qualified teachers did not predict larger achievement gaps between high-SES and low-SES students cross-nationally. These analyses provide empirical, cross-national evidence of the importance of investing in teacher quality for improving national achievement. National policies and practices related to improving teacher quality appear to be a promising area for future research to identify how other countries have achieved both excellence and equity in student achievement. (Akiba, LeTendre, & Scribner, 2007 p.369)

As such, the article is discussing equity in education with an explanans of SES, but this is connected also to differences in teacher quality. We have now presented at least three different ways of gathering the explanans in the articles – SES and family composition, SES and instructional environment, and SES and teacher quality. The last example we will present is an article trying to find explanans in gathering SES with resources and institutional settings. The article is presented in *Economics of Education Review*, and the study is examining seven East European transition countries by estimating educational production functions (Ammermüller, Heijke, & Wössmann, 2005). The results show substantial effects of student background on educational performance, but it can also show that the impact of resources and institutional setting is much lower.

The articles presented under socio-economic matters all show similarities in trying to explain differences in mathematics performance from differences in students' SES. This is handled in different ways within the articles. In summary, it can be said that all the articles have trouble finding strong correlations when going outside the SES explanation and all of them similarly conclude that SES alone is the most evident indicator in explaining differences. As such it can be said that the most used explanantia for explaining differences in performances are based on student characteristics, even though educational measures and contextual circumstances are used as complements in the argumentation.

To sum up

We have here showed how we have synthesized research publications dealing with equity problematics. The conclusions are a bit different considering the different issues.

- *Gender matters*: The gender gaps open up going in different discussion directions, depending on which indicators are in use for explaining the gaps within different contexts, both in the contexts investigated – but also concerning which contexts the results are presented in terms of focuses in the journals.
- *Socio-economic matters*: The gaps visible in relation to differences in SES are also discussed in relation to different supporting indicators. The composition found in our material is basically SES connected to family composition, SES connected to instructional environment, SES connected to teachers' quality and finally, SES connected to resources and institutional settings.

Randomized inquiries

Under the heading of randomized problematics we include studies on student performances combined with a wide set of different explanans. When analyzing this group we have first included the articles discussing mathematics performance (presented in the table below) but also the studies not explicitly discussing performance in mathematics. As such, this section is discussed both in relation to the PISA study and the categorization of directives problematics, and also in itself. Therefore, some similarities can be seen, but the differences might be more apparent too. Most of the studies could also be grouped under the headings of equity or efficiency, but this has not been the case since they are not explicit on these claims. The articles explicitly investigating mathematics performance in this section are rather few (N:5), and because of that the discussion should not be taken as absolutely true. For more ample discussion we have also included articles in the analysis that are not specific for mathematics achievement. In this section, further analysis has to be done for guaranteeing validity and reliability. Table 17 presents the 5 articles discussing mathematics performance without being categorized as either dealing with the efficiency or equity problematics:

Table 17. An overview of TIMSS research dealing with the randomized problematics.

| Reference | Explanandum | Explanans | Conclusion | Comment |
|-----------|---|---|---|---|
| 2002A | Math & Science performance + curriculum (nations) | Differences between applying and knowing | In relation to TIMSS – some did better on knowing | Not clear in argumentation |
| 2086A | Performance (math + science) + countries + over time changes + gender | Gender + differences in different areas | Gender gaps are complex, and even more over time. Locating male and female areas within TIMSS over time | The article is rather descriptive |
| 2171A | Performance (math) + differences in sub-skills | They create clusters of countries and saying that more in depth analysis has to be done concerning effects of culture and educational systems | The article shows differences in sub-skills | Not really relevant, only investigating sub-skills for creating clusters of countries |
| 2189A | Performance (math + science) + differences on item level + grade 7 | Age | The article indicate that not all TIMSS items have resulted in a higher mean | More than 40 countries Not really |

| Reference | Explanandum | Explanans | Conclusion | Comment |
|-----------|--|--|---|----------------------------------|
| | or 8 + countries | | score at the upper level | relevant Descriptive |
| 2204A | Performance (math + science) + knowledge at each proficiency level | Terminology and context, facts versus principles, and ability to communicate knowledge | Showing the strength and weakness in descriptions of the different benchmarked proficiency levels | Descriptive and subject-specific |

Most studies are what we in this systematic review call *gap studies*, which try to explain why differences appear. The sample we have presented above contains rather few articles. Only five articles are categorized for having a special interest in investigating mathematic performance without explicitly focusing on either efficiency or equity. What is shown from the assembly above is the fact that the articles are rather wide in scope, aims, and also results. The scope of the studies stretches from a wide range of interests and different methodologies. As an example, the study made by Nixon & Barth (2014), presented in *School Science and Mathematics*, is rather inventive. In the study four classroom teachers categorized samples of TIMSS test items by cognitive domains of knowing and applying by using definitions provided internally by the TIMSS 2011 Assessment Frameworks. Items of different cognitive domains could then be analyzed separately. After this the performances on the item level was compared between six nations. As a result the study could show that in some nations there was no significant difference among the students between knowing and applying, but in some other nations the students had the knowledge but not the abilities to apply them. With this as the most relevant result, the authors suggest that policy makers, educators and citizens should be cautious when interpreting the results of TIMSS as visualized in rank tables.

The second study above, presented in *Studies in Educational Evaluation* by Neuschmidt, Barth & Hastedt (2008), is much more traditional in its implementation of research interest. Basically, the study tries to locate differences over time concerning the gender gaps in mathematics and science. This is done through comparing the results in TIMSS 1995 and TIMSS 2003. By using a regression approach to compare the trend data, the findings indicate no major changes for mathematics. Concerning science, however, some indications show that the gender gap may be closing, especially in the previously male dominated areas of chemistry and physics. This kind of study has, in its methodological approach and its aims, a lot of similarities with studies performed with gender as explanans in both the PISA study and the TIMSS study above when focusing efficiency and equity. The article is however rather descriptive, and due to that the theoretical framing is not associated to either efficiency or equity.

The third study, presented in *American Educational Research Journal* by Tatsuoka, Curtis, & Tatsuoka (2004), might have been categorized as equity or efficiency within “*the black box*” matters if they were more explicit on the problematics. Instead the study used a diagnostic testing approach to compare the mathematics achievement across a sample of 20 countries. By using a rule-space method student mastery was measured on 23 specific content knowledge and processing sub-skills. As a result, the authors could find patterns of sub-skill achievement that differed between nations. In the study criticism is formulated on methodological issues concerning how to interpret results in ILSA. It is expressed:

Factor analysis, cluster analysis, and traditional latent class models produce factors, clusters, and classes, but they are exploratory methods that merely group observed responses into similar classes or patterns. For this reason, they may produce solutions with no clear interpretation of the resulting groups of items or respondents. Ideally, diagnostic analyses of test results should be descriptive and objective, uniquely expressing an individual's state of knowledge, which must be free from ambiguous interpretations. To achieve these goals, we need a new methodology that transforms unobservable knowledge and subskill variables into observable variables. (Tatsuoka, Curtis, & Tatsuoka, 2004, p.904)

The model the authors therefore used for achieving their goals was the rule-space method. With this, a re-analysis of TIMSS data could reveal some new insights. The findings are then presented:

Our results show that high-achieving countries in the eighth-grade TIMSS- 99 mathematics assessment attained their level of performance in different ways. For example, Singapore students obtained the top performance on TIMSS mainly by showing excellence in reading and computational skills. Japanese students demonstrated excellent higher level thinking skills, while Belgian students achieved high scores through strength in fractions and proportional reasoning skills. Hong Kong and Korean students show relatively balanced knowledge and processing skills. In contrast, students in the industrialized countries that were not grouped into the highest achieving cluster (see Figure 8) tend to show weaker scores in these higher level mathematical thinking skills. These industrialized countries in lower-achieving sub-groups include Russia, Italy (in the second cluster), Canada, Australia, Finland, England, United States (in the third cluster), and Israel (in the fourth cluster). These higher level skills are extremely important for students to master in order to succeed at study or employment in science and technology fields. (Tatsuoka, Curtis, & Tatsuoka, 2004, p.922)

The study presented is an example of studies that are using TIMSS data for re-analyses or reorganization of the data for answering somewhat different questions than the ones asked within the TIMSS tests. We have found examples of this in both the PISA and TIMSS study; therefore, it can be said that ILSA is leading to both theoretical and methodological inventions, but at the same time it must be said that these studies are rather few.

The fourth article, by Wang & Zhu (2003) published in *School Science and Mathematics*, is a study trying to locate differences on item level dependent on which grade the students took the test. A peculiarity in TIMSS is that the students take the test in either grade seven or grade eight. The reasons for this is dependent upon national contexts, yet without going too deep into this inconsistency we can conclude that the authors of the article perform a comparison of 40 countries in order to sort out how this affects the item level results. The research findings indicate that not all TIMSS items have resulted in a higher mean score at the upper grade level. The study is trying to find out differences in performance by adding the explanans of the students' grade level, but basically they found out that this was not the most evident predictor of the results perceived.

The fifth and the last study in the table above is rather inventive. The author, Kelly (2002), set up two 10-member panels of subject-matter specialists for describing what they perceived as student proficiency at each level described in the TIMSS test. By that the author means that "world class" achievement could be defined. As a result the following is described as being the proficiencies needed for mathematics:

In mathematics, three factors appear to distinguish performance at the four benchmarks: the mathematical operation required the complexity of the numbers or number system, and the problem situation. For example, at fourth grade, students scoring at the lower end of the scale demonstrate facility with whole numbers in simple problem situations, while their peers scoring at higher levels on the scale can use basic operations on whole numbers and can solve multistep word problems. Students scoring at the Top 10% benchmark demonstrate an understanding of fractions and decimals and can perform simple division. At eighth grade, students scoring at the lower end of the scale demonstrate an understanding of fractions and decimals and perform basic operations on decimal numbers, while students scoring at higher levels can perform basic operations on fractions, locate and use data in charts and graphs to solve problems, and have a grasp of beginning algebra. Students scoring at the Top 10% benchmark demonstrate that they can "bring things together." They organize information in problem-solving situations and apply relationships to solve problems. (Kelly, 2002, pp.51-52)

What becomes obvious is that the proficiency levels described in the TIMSS test is analyzed and used as a stepping stone for immersing the discussion on what proficiencies students need for scoring high on the test. In this the explanandum is still to describe differences in performance, but the explanans is placed on different proficiencies the students have. We cannot quite see how this will help the authors explain why the differences appear. Instead, the study can be seen as highly related to policy and professionally relevant for training students in crucial parts of mathematics.

Besides these studies presented in the table above there are some more studies that could be categorized within the randomized problematics, but these studies are not focusing on mathematics. Instead science is in focus. Just in short, it can be said that they also place the explanandum within trying to explain why differences in results appear and as such are gap studies, but when it comes to explanans they show a variety. In this

student characteristics such as home language (Hauger & Sireci, 2008), attitudes (Papanastasiou & Papanastasiou, 2004) or academic self-concept (Yu, 2012); educational measures such as teachers' competencies (Vlaardingerbroek & Taylor, 2003) or teachers' conception of math (Philippou & Christou, 1999); or contextual circumstances like poverty (Wiseman, 2012), cultural differences (Kjaernsli & Lie, 2008) or violence (Akiba, LeTendre, Baker, & Goesling, 2002) are used for explaining the differences in performance.

On TIMSS research compared to PISA research

As mentioned before, synthesizing research is at the core in a systematic research review, being aggregative or configurative or as in our case argumentative. However, as shown this is highly problematic, partly due to the research field, partly due to the task to synthesize research for translation into educational practice of different kinds. In Chapter 5 this is discussed more in detail.

What can be said anyhow is that when comparing PISA and TIMSS research some similarities and differences can be seen. The most obvious similarity is that both branches of ILSA are wide in scope, interest and levels of focus. There is a similarity also explicit in that both equity and efficiency problematics are evident in both PISA and TIMSS research. A difference is noted when it comes to the third problematics – in PISA they are called direction problematics, which try to deal with explananda in career expectations or interest or attitudes combined with taxonomic groups. This specific problematics is more or less absent in the TIMSS research. Instead we have grouped studies within the TIMSS research under the heading of randomized problematics. Even though some of the articles under this heading could be categorized under the same explanandum as direction problematics, these are sparse and also not consistent with the much more focused problematics evident in PISA research. What this difference stands for could be up for debate, but at least one possible explanation has to do with the OECD being an outspoken economic association, which the IEA is not. It could therefore be possible that PISA as an assessment is more focused and leading to more focused debates on issues related to future careers and employment. If this holds true, it might be possible to discuss differences between PISA and TIMSS research in terms of TIMSS research being freer and more inventive than PISA research when it comes to scope and content, but especially in adding various explanans on differences in performance.

The explanantia are similar in both PISA and TIMSS research when it comes to the three sets we have observed – student characteristics, educational measures and contextual circumstances. These different sets are however dealt with in different ways and different scopes come in focus depending on if the study is presented by data from PISA or TIMSS.

A concluding remark after conducting a comparison between PISA and TIMSS research can very well be formulated in terms of ILSA research seeming to have similarities, but there are also disparities between the two branches studied so far. A very well grounded question so far can be if it is possible to talk of one single ILSA research. Or, is maybe more correct to talk about different traditions related to what data is used?

REVIEW OF THE CIVED/ICCS STUDIES

In this chapter we present a review of research based on the IEA programs CIVED and ICCS. This chapter is performed in a slightly different way when compared with the two previous chapters. This has to do with two things: first, CIVED and ICCS are completely different studies compared to PISA and TIMSS when it comes to scopes and aims. Second, mathematics is not measured in CIVED and ICCS, so that part is not relevant for comparisons. Similarities are that the filtering of articles has been conducted in the same way, and in this chapter we also deal with explanandum and explanans for synthesizing the content of the articles.

Field review and filtration process

The search process concerning CIVED/ICCS has been conducted in the same manner as the PISA and TIMSS study but with one important exception. When conducting a search combining CIVED with ICCS we noted that most of the ICCS studies were excluded. To compensate for this we made a compensatory search just including ICCS and then manually combined these two searches. The process was time consuming, but we concluded that this was the only way to deal with the shortcomings of Discovery. In the end that many articles appeared in the exclusively ICCS search, but to present a full coverage of CIVED and ICCS this time consuming handling was necessary. Also in this search the drawbacks concerning research presented in Swedish were evident, and in the same manner as the PISA and TIMSS study we also are missing publications not explicitly including education in their abstracts.

Inclusion and exclusion of articles

There are some major differences in studying CIVED/ICCS research compared to the studies based on PISA and TIMSS. The most evident is that the total number of articles that we collected with our inclusion and exclusion criteria is far less than the previous searches. This is not that surprising for us as researchers involved in research about ILSA, but it is surprising that the differences were that large. An explanation of this might be the fact that research within this field is perhaps not that explicit in their relationship to the ILSA per se; rather, the studies discussing democracy, democratic participation and civic knowledge are wider in their scope on these issues and as a result do not combine the ILSA and education explicitly in their abstract. For us, this lead to having to grasp this field of research in a way other than the studies of PISA and TIMSS research. This will be discussed further in the chapter. Below follows an overview of the filtration process in different steps and criteria used in each step.

Table 18. Filtration of CIVED/ICCS articles in different steps by Discovery and manual processes.

| The search process of Discovery | |
|---|-----------------------|
| Step I:a | |
| Discovery search: "International Civic and Citizenship Education Study" OR "Civic Education Study" OR "CIVED" OR "ICCS" AND "education" | |
| Limited to AB Abstract and Search mode: Set for Boolean/phrase | |
| 947 articles | |
| Step II:a | |
| Limit to "Peer Reviewed" | |
| 658 articles included | 289 articles excluded |
| Step III:a | |
| Limit the publication date to 1998-2014 | |
| 653 articles included | 5 articles excluded |
| Step IV:a | |

| The search process of Discovery | |
|---|-----------------------|
| Limit to papers in academic journals | |
| 647 articles included | 6 articles excluded |
| Step V:a | |
| Limit by subject ¹⁸ | |
| 406 articles included | 241 articles excluded |
| Step VI:a | |
| Limit to English | |
| 384 articles included | 22 articles excluded |
| Step VII:a | |
| Limit by publication ¹⁹ | |
| 215 articles included | 169 articles excluded |
| Step VIII:a | |
| Removal of duplicates by Discovery | |
| 82 articles included | 133 articles excluded |
| Complementary search (due to that articles concerning ICCS was neglected in the first search) | |
| Discovery search: "International Civic and Citizenship Education Study" OR "ICCS" AND "education" | |
| Limited to AB Abstract and Search mode: Set for Boolean/phrase | |
| 745 articles | |
| Step II:b | |
| Limit to "Peer Reviewed" | |
| 507 articles included | 238 articles excluded |
| Step III:b | |
| Limit the publication date to 2008-2014 | |
| 386 articles included | 121 articles excluded |
| Step IV:b | |
| Limit to papers in academic journals | |
| 385 articles included | 1 articles excluded |

¹⁸ Note that the articles are classified more than once within Discovery. Subjects chosen: civics -- study & teaching (n:104), citizenship (n:65), civics (n:58), citizenship education (n:55), civic education (n:46), democracy (n:45), education (n:39), foreign countries (n:35), political participation (n:29), political socialization (n:29), social studies (n:26), social participation (n:23), academic achievement (n:22), citizenship -- study & teaching (n:21), secondary education (n:21), comparative analysis (n:17), education & state (n:17), higher education (n:17), political science (n:17), student attitudes (n:17), curricula (courses of study)(n: 15), citizen participation (n:14), adolescents (n:13), elementary secondary education (n:13), human rights (n:13), knowledge level (n:13), research (n:13), teaching (n:13), classroom environment (n:12), social justice (n:12), students -- attitudes (n:12), study & teaching (n:12), educational sociology (n:11), political attitudes (n:11), social sciences (n:11), social sciences -- study & teaching (n:11), teenagers (n:11), education -- united states (n:10), comparative studies(n:9), cultural pluralism (n:9), democratic values (n:9), educational psychology (n:9), teaching methods (n:9), universities & colleges (n:9), comparative education (n:8), education & politics (n:8), international education (n:8), learning (n:8), nationalism (n:8), political psychology (n:8).

¹⁹ The following journals have been chosen (note that some of the journals can appear more than once due to slightly similar spellings or presentation. Discovery does not sort these out): Social Studies (n:15), Applied Developmental Science (n:13), Educational Psychology (n:12), Oxford Review of Education (n:12), International Journal of Educational Research (n:10), European Educational Research Journal (n:8), Educational Research (n:7), Reading Teacher (n:7), British Journal of Educational Studies (n:6), Comparative Education Review (n:6), Comparative Education (n:5), Political Psychology (n:5), Social Education (n:5), American Behavioral Scientist (n:4), British Educational Research Journal (n:4), International Sociology (n:4), Journal of Educational Research (n:4), Political Behavior (n:4), School Effectiveness & School Improvement (n:4), Studies in Higher Education (n:4), American Psychologist (n:3), Change (n:3), Compare: a Journal of Comparative Education (n:3), Democratization (n:3), Education & Society (n:3), Elementary School Journal (n:3), International Journal of Testing (n:3), Japan Forum (n:3), Journal of Ethnic & Migration Studies (n:3), Journal of Social Issues (n:3), Journal of Youth & Adolescence (n:3), Public Organization Review (n:3), Research Papers in Education (n:3), Social Science Journal (n:3), Systems Research & Behavioral Science (n:3), Teaching Sociology (n:3), European Sociological Review (n:2), International Journal of Progressive Education (n:2), Social Philosophy & Policy (n:2), Teacher Librarian (n:2), Young (n:2), African Journal of Political Science & International Relations (n:1), American Journal of Orthopsychiatry (n:1), Business & Management Research (n:1), Communication Education (n:1), Great Plains Research (n:1), Journal of Public Affairs Education (n:1), Learning Languages (n:1), Ohio Social Studies Review (n:1), Pacific-Asian Education Journal (n:1).

| The search process of Discovery | |
|---|---|
| Step V:b | |
| Limit by subject ²⁰ | |
| 182 articles included | 203 articles excluded |
| Step VI:b | |
| Limit to English | |
| 174 articles included | 8 articles excluded |
| Step VII:b | |
| Limit by publication ²¹ | |
| 74 articles included | 100 articles excluded |
| Step VIII:b | |
| Removal of duplicates by Discovery | |
| 30 articles included | 44 articles excluded |
| Manual process of including and excluding articles (this process is made explicit in Appendix 1 | |
| Step IX | |
| Manual removal of articles that are duplicates in the second search in relation to the first search | |
| 16 articles included | 14 articles excluded |
| Manual removal of duplicates based on abstracts | |
| 93 articles included | 5 duplicates removed manually |
| Step X | |
| Manual removal of newsletter based on abstracts ²² | |
| 93 articles included | 0 newsletter/grey literature removed manually |

²⁰ The following subjects are chosen: Citizenship (N:50), Citizenship education (N:28), Civics – study & teaching (N:25), Research-evaluation (N:25), Citizenship – study & teaching (N:21), Foreign countries (N:20), Civics (N:19), Research (N:19), Education (N:16), Psychometrics (N:16), Higher education (N:14), Political participation (N:12), World citizenship (N:12), Evaluation (N:11), Adolescents (N:10), Civic education (N:10), Cross-cultural communication (N:10), Comparative analysis (N:9), Curriculum planning (N:9), Democracy (N:9), Multicultural education (N:9), Student attitudes (N:9), Teaching methods (N:9)

And the following is not chosen: Data analysis – software (N:18), Statistics (N:37), Inter-observer reliability (N:36), Questionnaires (N:35), Reliability (N:35), Research methodology evaluation (N:29), Correlation (statistics) (N:28), Descriptive statistics (N:24), Confidence intervals (N:22), Research – finance (N:20), Analysis of variance (N:18), Data analysis (N:17), Research- methodology (N:16), Validity (N:16), Chi-squared test (N:15), Scale analysis (psychology) (N:15), Quality of life (N:14), t-test (Statistics) (N:14), Factor analysis (N:13), Physical education (N:12), Longitudinal method (N:11), Cross-sectional method (N:10), Public health (N:10), Educational psychology (N:9), Observation (scientific method) (N:9), Patients (N:9), Sampling (statistics) (N:9)

²¹ The following journals are chosen: Educational Psychology (N:12), Oxford Review of Education (N:7), Teachers College Record (N:6), Educational Research (N:5), British Educational Research Journal (N:4), Studies in Higher Education (N:4), Applied Developmental Science (N:3), Journal of School Health (N:3), Journal of Studies in International Education (N:3), Public Organizer Review (N:3), Quality Assurance in Education: an International Perspective (N:3), Citizenship Studies (N:2), Education, Citizenship & Social Justice (N:2), Education, Citizenship and Social Justice (N:2), European Educational Research Journal (N:2), International Journal of Progressive Education (N:2), Irish Educational Studies (N:2), Journal of Curriculum Studies (N:2), Journal of Education (N:2), School Effectiveness & School Improvement (N:2), Annual Review of Education, Communication & Language Sciences (N:1), International School Journal (N:1), Journal of Social Sciences (N:1)

And the following is not chosen: Journal of Interprofessional care (N:6), Intercultural Education (N:5), Disability & Rehabilitation (N:4), Journal of Social Issues (N:4), Aging & Mental Health (N:3), International Journal of Geriatric Psychiatry (N:3), Journal of Continuing Education in the Health Profession (N: 3), Journal of the American Geriatrics Society (N:3), Medical Teacher (N:3), BMC Health Services Research (N:2), European Journal of Cardiovascular Nursing (N:2), Home Health Care Management & Practice (N:2), International Journal of Behavioral Nutrition & Physical Activity (N:2), International Journal of Mental Health & Addiction (N:2), Journal of Clinical Nursing (N:2), Language & Intercultural Communication (N:2), Journal of Epidemiology & Community Health (N:2), Journal of Evaluation in Clinical Practice (N:2), Journal of Psychoactive Drugs (N:2), Respiration (N:2), Young (N:2), Canadian Issues/Themes Canadiens (N:1), Iberica (N:1), Intercultural Communication Studies (N:1), International Journal of Intercultural Relations, Journal of Applied Statistics (N:1), Kinesiology Slovenica (N:1)

²² In the process of reading the abstracts, articles have been removed for being what we call newsletters/grey literature. These are publications with mainly just an informative content about CIVED and/or ICCS. Often they are written by CIVED and/or ICCS government officials for informative purposes.

| | | | |
|---|-----------------------|---|--------------|
| Step XI | | | |
| Manual removal of editorial text based on abstracts | | | |
| 92 articles included | | 1 Editorial removed manually | |
| Step XII | | | |
| Manual removal of misplaced articles based on abstracts ²³ | | | |
| 54 articles included | | 38 articles removed | |
| Step XIII | | | |
| Manual classification of the abstracts in: | | | |
| A: Articles using CIVED and/or ICCS data for answering research questions | | | |
| B: Articles discussing CIVED and/or ICCS in policy reasons | | | |
| C: Articles discussing CIVED and/or ICCS and methodological implications | | | |
| A: 43 | | B: 5 | C: 6 |
| Step XIV | | | |
| Manual re-classification after reading available articles | | | |
| A: 40 articles included | B: 5 articles removed | C: 8 articles removed | Misplaced: 1 |
| Step XV | | | |
| Excluded due to availability | | | |
| 39 articles available in full text | | 1 articles not available in full text ²⁴ | |
| Step XVI | | | |
| Manual classification after reading the articles in: | | | |
| A1: Comparing/discussing CIVED and/or ICCS in relation to two or more countries | | | |
| A2: Only discussing CIVED and/or ICCS in relation to one country | | | |
| A1. 23 articles included | | A2: 16 articles excluded | |

After manual removals of misplaced articles and duplicates, we ended up with 54 articles. Of these, were 43 consistent with our search criteria while the others were excluded. Of the removed 5 were discussing CIVED/ICCS in policy terms and 6 were dealing with methodological issues. After that we excluded articles only discussing CIVED/ICCS in relation to one country. That left us with 23 articles discussing two or more countries. These are the articles we then have analyzed and synthesized.

Screening and coding selected publications

After the initial screening of research, we made further analyses to filter the selection through different steps.

In the coding process we described every publication in the different aspects according to our design. In the table below we describe the publications over the different aspects we used. We will now present the aspects and what distinctions are used to describe the studies.

Comparable with the previous studies, these studies present themselves as being initiated by the researchers/authors themselves. Also in this case the filtration criteria for all studies are based on comparisons between different countries with either CIVED or ICCS as its common point for developing comparisons. In this screening process it becomes evident for us that research discussing issues under scrutiny in the CIVED/ICCS tests, such as democracy participation and civic knowledge, is a much larger field than the studies included in our Discovery search with rather sharp and narrow inclusion and exclusion criteria. There are a lot more articles discussing these issues with references to CIVED/ICCS than the articles included in our

²³ These articles contain “CIVED” and/or “ICCS” AND “education” in their abstracts, but that is just for discussing these phenomena in relation to what is actually investigated.

selection. The same can be said to be true also for the PISA and TIMSS study, but we note that the case is clearer in the field of CIVED/ICCS inquiries.

Second, we note that in our selection none of the articles discuss CIVED/ICCS in relation to other ILSA. Instead, the vast majority of articles (N:20) only discuss within exclusively CIVED/ICCS data, or a minor part (N:3) in relation to other kinds of statistical data. Concerning what kind of level the research question is based on, the result of our mapping is more or less similar to the previous mappings. The CIVED/ICCS research is also primarily interested in individual factors (N:11), followed by organizational (N:8), system (N:6), but with some subject-specific (N:2) studies made as well. In relation to the last group, a rather interesting pattern emerged: PISA, organized by the OECD, lacked subject-specific research, but the IEA studies have this content within their research, following the specific ILSA presented in academic journals. In the case of TIMSS, mathematics and science are the prime subjects, and in CIVED/ICCS the subject is civics.

Once again, by using the distinction between intensive and extensive research programs (Sayers, 2000) we note that what is done in most of the studies is a search for regularities or significant patterns concerning taxonomic groups and their relative achievement.

Table 19. Describing selected CIVED/ICCS publications over different aspects. (N:23)

| Aspects | Alternatives | Distribution |
|---|--|---------------------|
| What to be inquired? | Student achievement | 16 |
| | School characters | 6 |
| | Education system characters | 1 |
| How to be inquired? | Formal (only CIVED/ICCS) | 20 |
| | Formal plus formal (TCIVED/ICCS + other LSA) | 0 |
| | Inside – outside (using CIVED/ICCS for inquiring other social phenomena) and substantial | 3 |
| | | |
| What research question level?* | Individual | 11 |
| | Organizational | 8 |
| | System | 6 |
| | Subject-specific | 2 |
| Type of account | Descriptive | 0 |
| | Explorative – hypothesis-generating | 15 |
| | Hypothesis-testing | 8 |
| What relevance – for whom? | Professional relevance | 6 |
| | Policy – administrative relevance | 14 |
| | Academic relevance | 5 |
| | Citizen relevance | 0 |
| | General information | 0 |
| What is the logic from research to practical relevance? | Directives – how to improve achievements? | 11 |
| | Insight – how to understand achievements? | 12 |
| | Not related | 0 |

Third, we looked for what kind of accounts, concerning knowledge production, the selected studies had. In this we found no descriptive articles at all. This is probably dependent on the rather narrow inclusion and exclusion criteria we had for our study. Within the total search we made in Discovery, we had several of these kinds of studies, but they were excluded due to our criteria. Instead, we note that explorative and hypothesis-generating accounts are the most evident (N:15), followed by hypothesis-testing (N:8). Before, in the PISA and TIMSS study, we concluded in that the major focus in explorative studies was rather surprising – the same can be said to be true concerning research based on CIVED/ICCS.

Fourth, just as in the PISA and TIMSS study (but not shown in the table below), there is a variation of knowledge objects in the studies under scrutiny. Often we also found an interest in taxonomic classifications of

groups in relation to school performance (in this case described in terms of democratic participation and civic knowledge) with an ambition to explain achievement gaps over these groups. This can be dealt with either in e.g. comparisons between countries, clusters of nations, SES or gender. This aspect will be presented more thoroughly when we describe the mapping of the research.

Fifth, when analyzing issues of relevance another pattern of interest emerges. Once again we found little research devoted to the academic community (N:5), but in this case an underestimation is probably attached to these findings since most of the research also includes a review in the publication. However, when we read the conclusions, issues of relevance are mostly addressed within political and administrative frameworks (N:14) and/or argued to be of professional relevance (N:6). To this is also added that most of the publications are assumed to provide new insights (N:12) or present directives (N:11) on how to increase performance on the issues measured in CIVED/ICCS.

With this coding we have given an overview of the publications presenting a research field whose variance is within sets of principles that give the field a rather homogeneous quality in one significant way – it is a research field with an ambition to identify significant patterns of regularities through correlations that derive from the statistical tools applied – normally within a discussion about achievement gaps. In this aspect the research field of CIVED/ICCS has many similarities with the research field illuminated within our PISA and TIMSS studies. The consequence of this is that we once again have to make a more detailed mapping of what kinds of knowledge is searched for and what kinds of understandings are presented.

Mapping explanandum and explanans in CIVED/ICCS studies

By explanandum we once again refer to what is to be explained. Just as before, this does not imply that such explanations are due to causal mechanisms, instead patterns or regularities are foregrounded in order to put forward conclusions. In the CIVED/ICCS research we once again find notions of “*gaps*.” These are discussed in the same manner: opportunity, learning or achievement gaps.

What is to be explained are often achievement gaps measured as different performances that are then correlated by pointing to various patterns of factors. We identified in relation to the CIVED/ICCS articles two different, but sometimes interrelated, research problematics:

- *Efficiency Problematics*: The studies categorized to deal with this problematics are studies involved in discussions on how to make schools more effective and how to perceive classrooms with a more positive attitude and a higher degree of openness, which is considered to be effective methods for increasing achievements and what kinds of artifacts to use for better performances. The categorization is also dependent on articles discussing differences across nations and between individual students in civic knowledge and why these appear.
- *Democracy Problematics*: The categorization is based on articles discussing CIVED/ICCS in relation to issues like participation and engagement in politics and attitudes among students concerning immigrants and human rights.

Explanans

As stated above in the PISA and TIMSS studies, explanandum and explanans are related to each other in various ways. We will once again try to present a set of overlapping principles. Below is our reasoning about kinds of explanans in relation to data (again we are trying to do this comparable to the prior studies on PISA and TIMSS).

- We have first the *internal explanans*, where different variables in the same dataset (in this case the various CIVED and ICCS tests) are related to each other – e.g. knowledge in civics and classroom openness.
- We have *external explanans*, where different datasets are combined — e.g. to capture attitudes towards immigrants over time. The point is that different measurements are related to each other in order to develop patterns or regularities.

- We have also sorted out one external explanans, called *taxonomic groups*, e.g. SES or gender that can be related to different other aspects e.g. attitudes towards immigrants or knowledge in civics.
- The logic or principles that relate to internal and external (also including taxonomic groups) are specified, tacitly or explicitly, through an abstract model of the *school as system* within a larger set of societal organization. In this the explanans is described within a larger set of describing differences in national contexts.
- The last explanans we have seen within the articles in our sample is an explanans derived from organizational principles, where *the school as a micro-cosmos* can form and organize the activities, e.g. performance and open classroom climate.

These different kinds of explanations can be combined in different ways in order to deal with the explanandum. In the same way as we could show in the PISA and TIMSS study, the research results are defined by explanandum in relation to explanans.

Synthesizing

Once more we will carry out a synthesis in terms of different explananda. We stated above that the focus in most of the CIVED/ICCS research has the same focus as PISA and TIMSS research in terms of the focus of different *gaps*. The syntheses will be presented over combinations of explanans and explanandum. Explananda can be formalized in two sets as follows:

A: Efficiency Problematics = School Performance or Knowledge Organization + Educational Measures + Taxonomic Groups

B: Democracy Problematics = School Performance + Attitudes or Taxonomic Groups

The efficiency set of explananda is most commonly trying to identify biases for different taxonomic groups defined by their differences and interpreted as indicators of differences in civic knowledge. These articles are trying to sort out what kind of methods, classroom climate or use of specific artefacts that is most effective for raising standards of civic knowledge. The democracy set is trying to sort out attitudinal differences affecting performances or differences related to different taxonomic groups.

What are then conceptualized as explanans to these different explanandum? In a broad sense we have identified the same sets as before:

1. Student characteristics
2. Educational measures
3. Contextual circumstances

In this we are reasoning in the same way as we have done in the previous PISA and TIMSS reviews.

Table 20. A structure of inquiries in CIVED/ICCS studies. Explananda and explanantia over publications.

| Explananda | Explanantia over publication numbers | | |
|-----------------------------|---|--|---|
| | Student characteristics | Education measures | Contextual circumstances |
| A Efficiency problematic | 3002A, 3003A, 3038A, 3021A, 3038A, 3063A | 3002A, 3003A, 3021A, 3023A, 3037A, 3038A, 3066A, 3069A | 3002A, 3021A, 3025A, 3037A, 3038A, 3072A, 4006A |
| C Democracy problematic | 3005A, 3007A, 3010A, 3011A, 3015A, 3018A, 3035A, 3065A, | 3005A, 3011A, 3018A, 3019A, 4008A, | 3005A, 3007A, 3010A, 3018A, 3019A, 3035A, |

Efficiency problematic research

Under the heading of efficiency problematics we can see some differences in research interest when it comes to scopes and aims. We have categorized these under the headlines of explicit efficiency matters and knowledge matters. Under each of the categorizations we will present some similarities.

Explicit efficiency matters

The matter of explicit efficiency is evident in our material, but there are only four articles explicitly stating this as a major focus in the article. Besides these, some of the other articles classified as belonging to other matters could be uplifted for making an argument on solutions for practitioners and policy makers in regards to how to increase civic knowledge with the help of proposed methods, processes or artefacts.

In a study by Isac, Maslowski, Creemers, & van der Werf (2014), published in *School Effectiveness and School Improvement*, the authors use an educational effectiveness approach to model the impact of student, school and educational system characteristics on several cognitive and non-cognitive student outcomes related to civic education. The result presented indicated that schools have a small influence on students' civic knowledge and hardly any impact on civic attitudes and intended civic behavior. Instead civic competences were mainly explained by individual student characteristics and out-of-school factors. One of the few factors making a difference in civic competences were related to stimulating a democratic classroom climate with free dialogue and critical debate, nurturing positive interpersonal relationships and creating opportunities for students to learn and practice democracy. The article by Isac, Maslowski, Creemers, & van der Werf had a forerunner in an article published a few years before in the same journal by nearly the same authors. In that article Isac, Maslowski, & van der Werf (2011) came to the same conclusion, but this first time they did not use ICCS as data, instead CIVED was used. The methodology was also comparable to the replicate. These findings can be said to be supported by the findings of Lin (2014), in an article published in *Applied Developmental Science*. By performing three-level hierarchal linear modeling Lin came to the conclusion that students' perception of an open classroom climate strongly related to higher civic knowledge scores in ICCS. In this case the relationship between open classroom climate and performance is also investigated in relation to SES. The study finds that an open classroom climate is supportive no matter the student's SES. In a last study classified as being conducted within the explicit efficiency matter, Janmaat (2011), in an article published in *International Sociology*, investigates whether civic competences among youngsters are linked to the social and ethnic composition of classrooms and if these links are influenced by the system of ability grouping. Janmaat concludes in that inequalities of civic competences across classrooms are relatively large in systems characterized by early selection on the basis of ability.

Within the few articles we have synthesized, we can conclude that there seems to be a unification of results highlighting the open classroom climate and abolition of grouping for increased civic performances.

Knowledge matters

Knowledge matters are more visible in our material than the explicit efficiency matters. There are a total of 9 articles dealing with this matter explicitly. The articles have, in the table above, been grouped due to explanans in student characteristics, educational measures or context matters. Several of them have explanans in more than one set. In some of these articles civic competences are discussed and how they might be understood. In an article by Hoskins, Barber, van Nijlen, & Villalba (2011), the complexity of the various influences affecting achievements in civic is discussed. The results indicate that a nation's years of democracy play both a positive and negative role on different aspects of civic competence, while citizenship education has better than expected consequences in some countries. As such, the last result is contradictory to what was found in some of the research within the explicit efficiency matters in that sense that they concluded that a school's influence on civic knowledge was rather minor. This is maybe not that surprising when considering the following expressed in their conclusion:

This article represents a first attempt to measure -using CIVED data - civic competence as a holistic concept combining cognitive (knowledge and skills) and affective (attitudes and values) dimensions together. The results of comparing European countries on this composite measure as well as the four domains of civic competence (Citizenship Values, Social Justice Values and Attitudes, Participatory Attitudes, and Cognitions about Democratic Institutions) offer insights, but they must also be considered exploratory in their nature and tentative in their interpretation. There are some caveats to the results. For instance, the CIVED data do not include measures of all the aspects of civic competence as conceptualized above. Additionally, the analyses were conducted with data from 14-year-old pupils only and may not represent the populations more generally in the countries studied. (Hoskins, Barber, van Nijlen, & Villalba, 2011, p.105)

What becomes evident is that when using different kinds of explanans, the results can appear to be different too. This does not mean that any of the articles are wrong in their conclusions; it just means that when using different explanans and performing research in different problematics one might end up with different and, when compared to other research, contradictory results.

Another study performed by Kennedy (2007) is trying to investigate what the students by themselves mean by the term of "*active citizenship*." The results show that students have rather sophisticated conceptions of citizenship responsibilities although their attitudes are gendered. As such, Kennedy offers a concept of what is actually meant, and in the next step it actually is measured in CIVED, with the students' own meaning of civics. The result also show that the youngsters seem committed to political obligations rather than social obligations, and they do not seem inclined to take advantage of their political rights or become involved in political activism. As such, the results can also be discussed in relation to what we below discuss in terms of democracy problematics.

Another study trying to discuss what knowledge actually is in relation to civics is performed by Maiello, Oser, & Biedermann (2003) and published in the *European Educational Research Journal*. The authors are trying to develop an alternative model to explain the likelihood of voting. The results indicate that home environment and school-related factors predict civic knowledge and skills equally well. This knowledge in civics is seen as a basis for democracy participation, and as such could also been discussed in relation to democracy problematics. Another angle on the same topic – how to discuss indicators for performance of knowledge in civics – is made by Alviar-Martin, Randall, Usher, & Engelhard (2008) when they discuss the confidence of teachers from four nations in addressing civics topics with their students. Results revealed that the confidence of teachers varied when they addressed civics topics. They could also show that the confidence varied dependent on the country of residence, and as such they conclude in that socio-based factors, school-based factors and public discourse illuminated effects of national context on teachers' confidence. Interesting to note is that the study does not put these findings in relation to students' actual performance on ILSA tests on civics.

In summary it can be said that the articles uplifted above (also including the articles on efficiency problematics) are all pretty close to the IEA conclusions presented in CIVED and ICCS. As such, the research is rather dependent on the data they use. But at the same time, it can be said that the research field of

CIVED/ICCS is much freer in comparison to the PISA and TIMSS research. This might be dependent on the fact that both PISA and TIMSS research are reliant upon performances of students on the test. Even though the assessments contain a lot of other information than students' results, these facts are to a much lesser extent used for research. In CIVED/ICCS research it can be said to be the opposite. A lot of the studies are rather uninterested in actual student performances and show much more attention to the data for explaining the results. Of course this has to do with the composition of the different assessments, but this assumption is not all true. Researchers within PISA and TIMSS are absolutely free to develop their own use of the data, but their use seems to be within tighter frames of these research fields. Considering what has been visualized above from the different articles, it is interesting that the greatest spread of results from CIVED/ICCS is that for fostering civic engagement: schools have to ensure open classroom climate for discussing issues, emphasizing the importance of the electoral process and encourage a participative school culture. These findings are very much evident in all the articles, but it is used to take discussions further on issues like what is knowledge, how is that perceived by students and teachers and how to implement better methods for fostering new generations of democratic citizens.

With this said we will now go on to discuss the other problematics seen in the articles – the democracy problematics.

Democracy problematics

Democracy problematics is also grouped under two headings – political participation/engagement matters and attitudes towards immigrants/human rights matters. The articles grouped under the respective heading have, to some extent, similarities and, to some another extent, disparities. A lot of the studies depend on taking for granted expectations fetched from other articles or the ILSA itself lifting alternative explanans, but nevertheless some differences in focus concerning explanans can be discussed.

Political participation/engagement matters

Articles grouped under this heading are primarily categorized due to the relation between students' knowledge and how this affects their actions in political participation and engagement. These aims and scopes are explained in the same manner as we discussed above: visualizing differences in student characteristics, educational measures and context matters. Within our selection of articles there is a similarity that participation has been considered as a crucial dimension of citizenship and that experiences within civil society are viewed as a relevant opportunity for developing personal and social resources essential for the survival and expansion of democracy. Additionally, this interest is basically taken as a strand for participation experiences in adolescence and seems to be a good predictor of political engagement during adult life.

One study that is trying to offer explanans on individual, school and country level is written by Quintelier & Hooghe (2013). From the hypothesis of that participatory, democratic climate is associated with civic and political engagement. The article uses data from 35 countries included in the ICCS 2009 to conclude that individual student perception of a participatory democratic climate — especially openness in classroom discussions at the individual level – is positively associated with intended political participation. In using the same data (ICCS 2009) Hooghe (together with a colleague) also have an article investigating the intention of electoral participation. The study is comparing 14-year-olds in 22 European countries, and the results indicate that the willingness to vote remains quite elevated among the students studied but with a clear gender division. Girls are more likely to state that they will vote while boys are more likely to see themselves as future election candidates. The study could also show that an open classroom climate contributes to the willingness to vote in future elections (Hooghe & Dassonneville, 2013).

In a study by Menezes (2003), participation as an activity is most perceived in organizations that provide enrichment activities (e.g. sports, music, computers), but students are also involved in a variety of voluntary activities in some civics-related organizations (e.g. Scouts, religious affiliations and environmental clubs) and in experiences within the school (e.g. student councils and school newspapers). The author shows that cross-national and cross-age variation are significant. Overall there seems to be a positive impact of the frequency of students' involvement on civic concepts, attitudes and engagement, but the results also reveal that more is not

necessarily better. The most relevant implication is shown for the development of educational citizenship projects in that “*action*” can be a powerful learning tool and that the quality of participation experiences is crucial.

In a study by Hoskins, Janmaat, & Villalba (2012), CIVED 1999 is used for examining factors associated with the ways in which young people learn positive attitudes towards participation in, and knowledge and skills about, democracy. In this the authors use the hypothesis that less formal learning, in this case understood as social participation in the Lave and Wenger²⁵ definition, leads to better participation. The analysis shows that learning through social participation, inside and outside school and in particular through meaning-making activities, shows a strong positive relationship with citizenship knowledge, skills and dispositions. As such, the authors' argue in terms of situated learning theory as most useful for describing civic learning, and also for the relevance and applicability in ILSA research as a whole.

Attitudes towards immigrants/human rights matters

Research concerned with attitudes towards immigrants and human rights matters is categorized in two different groups, but they are interlinked and can be discussed in equal terms. There is a difference in relation to the same problematics above as grouped into political participation/engagement matters in the respect that the latter is more concerned with students' knowledge in relation to action and the former is more concerned with the students' knowledge related to their attitudes. The matter is dealt with in a separate way, but we can normally see that the explanans have the three sets we discussed previously: student characteristics, educational measures and contextual matters. The studies interested in student attitudes and human rights are primarily highlighting this from the strand that many countries witness the electoral success of anti-immigrant parties. As such, these articles are investigating how to abolish xenophobia, ethnocentrism, negative attitudes towards immigrants and the rejection of a multicultural society.

In an article by Elchardus, Franck, De Groof, & Kavadias, (2013) explanans is offered on student characteristics and contextual matters on students' acceptance of a multicultural society. This is done with an examination of the impact of private, quasi-market versus public steering of educational systems on European youngsters' attitudes towards immigrants. In using ICCS 2009 they can, in their results, visualize that quasi-market systems are observed to lead to less support for immigrants' rights, and this is largely due to higher concentrations of immigrant children in low SES schools in such systems. In the article it is stated that these characteristics of the educational system explains about half of the cross-national variation in attitudes towards immigrants among the 21 countries observed. In another study – also with an interest in adolescents' attitudes towards immigrants and how this is developed in tandem with the sense of national identity – Barber, Fennelly, & Torney-Purta (2013) use another set of data – CIVED 1999 – for concluding similar results. The attitudes were combined with national indicators of citizenship policies and demographics in a multilevel analysis. The results indicate that high levels of protective nationalism were associated with negative attitudes towards immigrants' rights in long-established democracies but not in newer ones. The result also showed that this relationship was stronger in religiously diverse countries. Further, adolescents in countries with more restrictive citizenship policies were less supportive of immigrants' rights although these policies did not moderate the extent to which attitudes towards immigrants were correlated with nationalism. As such, this article visualizes that attention to the national context as an importance explanans of social attitudes.

From the discussion above, a widespread belief can be seen – that ethnically mixed schools contribute to inter-ethnic tolerance and community cohesion. An article taking this belief as a stepping stone is written by Janmaat (2012) in comparing various individual and classroom variables in Sweden, Germany and England. The study finds a positive effect of classroom diversity on ethnic tolerance in Sweden and Germany, but it does

²⁵ Situated learning was proposed by Jean Lave and Etienne Wenger as a model of learning in a community of practice. At its simplest, situated learning is learning that takes place in the same context in which it is applied. Learning should not be viewed as simply the transmission of abstract and de-contextualized knowledge from one individual to another but as a social process whereby knowledge is co-constructed; they suggest that such learning is situated in a specific context and embedded within a particular social and physical environment.

not find an effect of tolerance in England. The study, based on these findings, concludes with the statement that country-specific factors are the one that shape tolerance, not only diversity in schools and classrooms.

Attitudes and tolerance towards immigrants is not the only focus that is grouped below the above heading. Also, human rights are a focus that is discussed in relation to democracy problematics. This is normally done from an understanding that human rights among young people forms a foundation for future support and practice of rights. In an article by Torney-Purta, Wilkenfeld, & Barber (2008) a linear modeling analysis examines student level predictors (e.g. gender and school experience) and country-level predictors (e.g. history of democracy) of rights-related knowledge and attitudes. The results show that countries with governments that pay attention to human rights in intergovernmental discourse have students who perform better on human rights knowledge items. In this, the result shows differences in gender. When looking at rights-related attitudes, the findings showed that students with more knowledge of human rights were more frequently engaged with international topics, and students in more open classes and school climates held stronger norms supporting social movement citizenship and had more positive attitudes towards immigrants' rights, and were also more politically efficacious. The same results are more or less similar in another of the articles in our selection – Torney-Purta & Barber (2011).

Summarizing CIVED/ICCS research

As we stated in the beginning of this chapter, there are some considerable differences between the research fields of CIVED/ICCS in relation to PISA and TIMSS research. What has been discussed is the fact that especially PISA but also TIMSS research seems to be more focused and framed by the sources of data. CIVED/ICCS research is as such freer in scope and aim. Noteworthy is that even if this might be true, it seems to depend on the original CIVED and ICCS assessments. As such, it can be stated that CIVED and ICCS are freer in their construction. It has also to do with the subject under scrutiny – civics as a knowledge object is rather “fuzzy” in its construction and has to be tackled in significantly different ways other than reading, mathematics or science.

In the synthesizing of CIVED/ICCS research we have found two primary problematics: one deals with efficiency and the other with democracy. As we stated above, these problematics overlap at times with each other, but there are some apparent differences. The most obvious has to do with that efficiency is directed to some kind of “action” while democracy is more focused on “attitudes” towards groups. In this is embedded a normative statement that negative attitude towards these groups are not able to combine with a high sense of democracy.

COMPARISONS OF REVIEWS OF PISA, TIMSS AND CIVED

The main objective of this part of the study is to analyze PISA's, TIMSS's and CIVED's/ICCS's impact on the international scientific community. Impact can be analyzed by different means. In this part we focus on the analysis of the results of research – in terms of analysis of articles published in international peer reviewed journals. This part has been divided into different sections in order to try and answer the following questions: i) what kinds of articles are most commonly disseminated (in terms of a classification of the articles in A, B and C)?, ii) what international research canon is created from the explicitly outspoken articles concerning PISA, TIMSS and CIVED/ICCS?, iii) what journals are in use for publishing research on these specific ILSA? and iv) what are the countries of affiliation of the researchers? In doing so we have been inspired by a methodological approach based on bibliometrics, which is mainly a quantitative analysis of publications for describing patterns within a specific research field.

The technique for bibliometrics was first introduced in the 1960s by Eugene Garfield, and after that many studies have used this methodological strategy for analysis of publications as an indirect measure of scientific results (Dominguez, Vieira, & Vidal, 2012). The limitations of these kinds of studies is well documented but there are also benefits, most evident is the fact that bibliometrics helps us with systemizing a large research field into a more explainable field in which similarities and differences might be visualized. For our help we have used Discovery as the search tool (for a description of this see above) with all its benefits and detractions, for the more detailed searches on citations we have used Scopus, and in some cases also JSTOR and Discovery.

In doing a systematic bibliometric review, some considerations have to be made. The most obvious is that it is time consuming: the data is massive and has to be stored and managed with exactitude. Because of this we have a rather extensive amount of appendixes attached to the report. This is also done because we believe that doing so will make it easier for researchers to control our results while also making studies of their own based on our empirical collection. As such, we offer an entrance or a “*fast-track*” into research based on ILSA.

Our own entrance into the research field of ILSA was to conduct a Discovery search on texts explicitly mentioning the various ILSA in combination with education within their available abstracts (for a more detailed explanation of this process see above in the analyses of PISA, TIMSS and CIVED/ICCS in Chapters 5, 6 and 7). This is done because our main focus is to study articles that are explicitly trying to communicate results based on ILSA within the research field of education. In our work we have found that it is definitely possible to describe the field with another entrance, such as ILSA in combination with e.g. “*workforce competences*,” “*employment*” or “*social exclusion*.” This is a rather sharp exclusion technique for describing the field, but as an entrance we found it useful. For compensating the exclusion of articles we conducted a review of the citations based on the explicit articles (we will now call the articles found in our basic search based on the abstract as the explicit articles) and how these were disseminated into implicit articles (those using or discussing ILSA without mentioning them in combination with education in their abstracts). Even though our data is comprehensive, we cannot state that we have created a full description of the field, but at least we can state that we have created a rather detailed and far reaching description of it. We have also collected the references in use within the explicit articles for trying to answer what resources the articles are based, but this study is incomplete and will take much more time for completion.

The filtration process of articles

As mentioned above, exactitude is of importance in conducting this kind of research so is also the presentation of the exclusion and inclusion of articles. Below we present how this is done in the Discovery search process. This is also presented more in detail in the chapters concerning the analysis of the different ILSA.

Table 21. Filtration process results in reviewing ILSA programs (Explained more in detail in Chapters 5, 6 &7).

| Steps in filtering | CIVED and ICCS | PISA | TIMSS | In total |
|--|----------------|------|-------|----------|
| 8) Hits in Discovery | 947 | 4406 | 3391 | 8744 |
| 4) In peer reviewed journals only | 658 | 2230 | 1879 | 4786 |
| 5) Publication date (differs for the separate ILSA) | 653 | 2225 | 1877 | |
| 6) Limit to academic journals | 647 | 2179 | 1829 | |
| 7) Limit by subject | 406 | 1811 | 1553 | |
| 8) Limit to English | 384 | 1662 | 1507 | |
| 9) Limit by publication | 215 | 849 | 766 | |
| 10) Removal of duplicates by Discovery (Note that in the case of CIVED and ICCS a complementary Discovery search has been carried out finding 16 more articles included in the study, see Chapter 7) | 82 (+16) | 357 | 317 | |
| 11) Removal of duplicates based on abstract | 93 | 333 | 299 | |
| 12) Manual removal of Newsletters | 93 | 258 | 265 | |
| 13) Manual removals of editorial articles | 92 | 248 | 259 | |
| 14) Manual removal of misplaced articles based on abstracts | 54 | 218 | 251 | |
| 15) Manual classification of the abstracts into A, B and C (see table 7 below for the result in total) | 43 | 138 | 182 | |
| 16) 14- Manual re-classification after reading available articles | 40 | 131 | 182 | 353 |
| 17) Available articles | 39 | 106 | 140 | |
| 18) Manual classification only in two or more countries (see Table 8 below for the result in total) | 23 | 59 | 53 | 135 |

Step 13 in the filtration process needs further explanation. When we manually read through the abstracts we performed a classification into three related but exclusive groups – called A, B and C. The classification is based on the content of the articles in which we saw different focuses on what the articles were trying to explore in relation to the ILSA. The first categorization – called A – contains articles where their abstracts use the data from ILSA to compare or explain students’ achievements, either in relation to one nation or to compare two or more nations. These are the articles we were most interested in for performing our analysis of explanandum and explanans. These articles were later, after reading through the full-text articles, manually classified into A1 or A2, meaning articles just discussing one single country or discussing two or more countries in comparison. Articles classified as B are articles more concerned with discussions on policy implications of the different ILSA. These articles were not analyzed in search for the explanandum or explanans, but they were analyzed and described as a part of the research field as such and in that respect we have performed analysis on them in terms of authors’ affiliation, where they are published and where they disseminate. The same is true for articles classified as C. These are the articles discussing ILSA in methodological terms. Below is a table presenting the filtering process under Step 13:

Table 22. Explaining the manual classification in filter 13.

| Manual classification in filter 13 | CIVED/ICCS | PISA | TIMSS |
|---|------------|------|-------|
| A: Using program data – to step 14 | 43 | 138 | 182 |
| B: Discussing programs in policy – excluded | 5 | 68 | 38 |
| C: Methodological discussions only - excluded | 8 | 12 | 31 |

After the initial classification of the articles, based on the abstract level, we downloaded the articles and read through them. In this we found that some of the articles were to be re-classified (shown in filtration Step 14). In this it become apparent that some of the articles were not available to us, so these articles we considered as “falling offs” (shown in filtration Step 15). After that we performed a more extensive analysis of the articles classified as A, and during that we performed the last step for selecting the articles. The articles were then regrouped into A1 and A2 – meaning articles either just discussing one single country or comparing two or more countries. In this the former were chosen for further mapping and coding.

Table 23. Explaining the manual classification in filter 16.

| Manual Classification in filter 16 | CIVED/ICCS | PISA | TIMSS |
|---|-------------------|-------------|--------------|
| A1: Comparing two or more countries | 23 | 59 | 53 |
| A2: Not comparing countries - excluded | 16 | 49 | 87 |

The filtration process described above ended up analyzing 59 articles using PISA, 53 articles using TIMSS and 23 articles using CIVED/ICCS for comparisons of two or more countries. These are the articles we then analyze in the following chapters for saying something of the explanandum and explanans of research stemming from ILSA. These articles were then codified with the help of a specific coding scheme we developed (appendix 21). This coding scheme was mainly constructed by us because we needed a classification tool for saying something on an overall level about what kind of research the different ILSA lead to in the scientific journals.

The coding of articles

In our classification on what the articles were trying to inquire, we found that the vast majority of the articles had a focus on student achievement, and that school characters and education systems received much less of a focus. That may not be surprising. However, in relation to that, the ILSA also make claims on school characters and education systems, so it might be illuminated as an area in which research has a potential for improvement. The interest in focus within the articles was mainly answered with the help of one individual ILSA, but some articles were trying to combine different ILSA or using other statistics for answering the questions in focus, yet a rather modest amount of articles just used one ILSA for their research. Questions can be asked in relation to this fact – in what ways is ILSA research isolated from the surrounding society? Why is ILSA research so isolated within one single survey? In what ways are this research based on a self-referential system? One way to answer these questions is to do what we are doing in this systematic review – trying to describe the field in terms of channels for publications, the canon of dominant articles, the dissemination of articles, the content of articles –and the part we do not do – what are the core references for the articles?

When it comes to the research question level, the results of our coding are more mixed. Many of the articles are trying to analyze differences in achievement based on individual differences; however, many of them also try to analyze at the system or organizational level for explaining differences in student achievements. This is not surprising in relation to how the ILSA are constructed, but more surprisingly is the fact that subject-specific considerations are rather few in comparison. Especially, the fact that PISA articles lack subject-specific interest is surprisingly.

When it comes to the account of the studies in the articles, it is surprisingly that not more of them are trying to conduct hypothesis-testing due to that the ILSA is rather well equipped for doing that kind of research. Instead, it is far more common to use the data in ILSA for explorative and hypothesis-generating research. There are also some articles conducting more descriptive explanations on what e.g. PISA is and what PISA measures, but due to our filtering of articles the vast majority of these have already been excluded. This type of account is mainly presented for policy makers with less explicit claims for professionals. Also, some of the articles make claims on academic relevance. Notably is the fact that none of the articles make claims on a general or citizen relevance. Finally, the articles are mainly concerned with making claims on what we call

insights on how to understand achievements and fewer of them make claims on directives of how to conduct education. This is rather surprising since most of the articles directed their relevance towards policy makers and professionals and since they maybe are not primarily interested in insights on how to understand achievements, which is maybe more interesting for the scientific community. Our systemic review is not the absolutely best way for trying to answer what this leads to, so other questions can be asked, like is this maybe an answer to why research using ILSA data is not that commonly used in policymaking and teaching and instead the grey-literature is used, especially among policy makers for making claims on education, and is this why organizations like McKinsey are offered a “window” in which they can act, based on results within ILSA? In other words, is ILSA research using a different narrative from the policy makers, making it hard for them to adopt the results without the last “chain” in the argumentation – making explicitly claims on directives for education?

Table 24. Coding results for selected publications from the ILSA programs PISA, TIMSS and CIVED. Numbers in total

| Aspects | Alternatives | Distribution PISA (n:59) | Distribution TIMSS (n:53) | Distribution CIVED (n:23) | Total |
|---|--|--------------------------|---------------------------|---------------------------|-------|
| What to be inquired? | Student achievement | 49 | 46 | 16 | 111 |
| | School characters | 2 | 2 | 6 | 10 |
| | Education system characters | 8 | 5 | 1 | 14 |
| How to be inquired? | Formal (only ILSA) | 46 | 43 | 20 | 109 |
| | Formal plus formal (ILSA in combination with other LSA) | 5 | 5 | 0 | 10 |
| | Inside – outside (using ILSA for inquiring other social phenomena) and substantial | 8 | 5 | 3 | 16 |
| | | | | | |
| What research question level? | Individual | 29 | 20 | 11 | 60 |
| | Organizational | 10 | 14 | 8 | 32 |
| | System | 20 | 16 | 6 | 42 |
| | Subject-specific | 0 | 6 | 2 | 8 |
| Type of account | Descriptive | 7 | 5 | 0 | 12 |
| | Explorative – hypothesis-generating | 44 | 23 | 15 | 82 |
| | Hypothesis-testing | 8 | 25 | 8 | 41 |
| What relevance – for whom? | Professional relevance | 17 | 12 | 6 | 35 |
| | Policy – administrative relevance | 40 | 39 | 14 | 93 |
| | Academic relevance | 3 | 2 | 0 | 10 |
| | Citizen relevance | 0 | 0 | 0 | 0 |
| | General information | 1 | 0 | 0 | 0 |
| What is the logic from research to practical relevance? | Directives – how to improve achievements? | 20 | 8 | 11 | 39 |
| | Insight – how to understand achievements? | 39 | 46 | 0 | 97 |
| | Not related | 1 | 0 | 12 | 1 |

The impact of PISA, TIMSS and CIVED/ICCS on academic journals

The aim of this part of the systemic review, as stated above, is to compensate for some of the limitations in relation to how we have selected the articles. Our selection of articles is based on articles explicitly referencing PISA, TIMSS or CIVED/ICCS along with education within their abstracts. There are far more articles referencing and discussing these ILSA than there are articles selected in our primarily Discovery search. One way to compensate for this lack is to go further and investigate the articles referencing our prime selection of articles. By doing so we state that we can give a more extensive insight of the PISA, TIMSS and CIVED/ICCS research available within academic journals. To help us in this part of the study we have used Scopus, JSTOR and Discovery as search tools for building a database consisting of the number of references connected to each of the articles in our primarily search, in what journal they are published (both the explicit and implicit articles), country of affiliation of the authors of the explicit articles, the most referenced articles within the explicit articles and the numbers of disseminated references by our classification in A, B and C.

In which journals are the explicit articles published?

We chose to describe this in a table below. We have just taken the most used journals for the different ILSA:

Table 25. The journals where most of the explicit articles are published.

| Journal | Number of PISA articles | Number of TIMSS articles | Number of CIVED and ICCS articles | Total number |
|--|-------------------------|--------------------------|-----------------------------------|--------------|
| Assessment in Education: Principles, Policy & Practice | 5 | 9 | 0 | 14 |
| Comparative Education | 10 | 2 | 1 | 13 |
| Comparative Education Review | 11 | 6 | 2 | 19 |
| Economics of Education Review | 8 | 4 | 0 | 12 |
| Education Economics | 7 | 2 | 0 | 9 |
| Education Journal | 8 | 0 | 0 | 8 |
| Educational Evaluation and Policy Analysis | 0 | 7 | 0 | 7 |
| Educational Research & Evaluation | 15 | 40 | 0 | 55 |
| Educational Studies in Mathematics | 1 | 7 | 0 | 8 |
| European Educational Research Journal | 20 | 0 | 7 | 27 |
| International Journal of Instructional Media | 0 | 25 | 0 | 25 |
| International Journal of Science Education | 15 | 8 | 0 | 23 |
| International Journal of Testing | 4 | 12 | 1 | 17 |
| Journal of Education Policy | 10 | 0 | 0 | 10 |
| Oxford Review of Education | 6 | 2 | 2 | 10 |
| Oxford Studies in Comparative Education | 12 | 0 | 0 | 12 |
| Scandinavian Journal of Educational Research | 11 | 0 | 0 | 11 |
| School Effectiveness & School Improvement | 6 | 0 | 2 | 8 |
| School Science & Mathematics | 0 | 12 | 0 | 12 |
| Studies in Educational Evaluation | 0 | 37 | 0 | 37 |

What becomes rather evident is that the different ILSA chose to publish in different research journals. Some of the journals can be seen as journals exclusively for discussions about one specific ILSA, as is the case with *Educational Evaluation and Policy Analysis*, *School Science & Mathematics* and *International Journal of*

Instructional Media only containing TIMSS articles or the *Journal of Education Policy* and the special issue of *Oxford Studies in Comparative Education* only containing PISA articles. Others are wider in scope and contain articles from more than one ILSA – such as *Comparative Education*, *Comparative Education Review*, *International Journal of Testing* and the *Oxford Review of Education*.

In systematizing these publications some rather peculiar circumstances appeared. One of the peculiarities is that some researchers seem to be rather productive in publishing articles in relation to ILSA. In this we have recognized a practice among some researchers to “build a career” on ILSA research. Another aspect of this is that several of the researchers within our scope are also researchers deeply involved in the different ILSA themselves. We are not criticizing this aspect of the research, we just conclude that the research field itself seems to be rather immature in this aspect and that the cross-section between research and what we call grey-zone seems to be rather evident.

What kinds of articles are most commonly disseminated?

Within our prime selection of articles containing explicit references to PISA, TIMSS, CIVED/ICCS and education within their abstracts we conducted a classification in three groups based on the major content of the articles. Articles classified as an A were articles using the data of the ILSA for doing either reanalyses, analyzes or comparisons for answering research questions. In our prime selection this group contained 138 PISA articles, 182 TIMSS articles and 43 CIVED/ICCS articles. These were then regrouped into A1 or A2 for articles either only discussing the results within one single country or articles discussing two or more countries for explaining a relevant question. The articles regrouped into A1 are the articles we analyze more properly in relation to explanandum and explanans.

Articles classified as B are articles with an outspoken interest in policy implications of PISA, TIMSS or CIVED/ICCS. The vast majority of the articles contain comparative ambitions with a terminology of borrowing and lending, globalization or internationalization of education. The group consists of 68 PISA articles, 38 TIMSS articles and 5 CIVED/ICCS articles. This says itself something about how the scientific community ranks the importance of the various ILSA.

The last classified group – C – is articles with a primary interest in methodological issues in relation to ILSA. This group consists of 12 PISA articles, 31 TIMSS articles and 8 CIVED/ICCS articles. Below follows a summarized table of the results but also an average of how many times the respective articles are cited. In this, Scopus is setting up some limitations because a number of the articles do have a loss in information that is also evident later on in this report: when we report the dissemination of articles, the loss is substantial, but despite this it is still possible to make some claims. Even though the loss is substantial, we describe this from a total number of 430 articles. The totality of articles citing our prime selection is presented in appendix.

Table 26. The total number of citations based on our classification.

| Classification | PISA | TIMSS | CIVED/ICCS | In total |
|----------------|--|--|---|---|
| A | 921 citations of 81 articles (11.37 citations/article) | 937 citations of 184 articles (5.09 citations/article) | 211 citations of 40 articles (5.28 citations/article) | 2069 citations of 305 articles (6.78 citations/article) |
| B | 493 citations of 38 articles (12.97 citations/article) | 266 citations of 34 articles (7.82 citations/article) | 27 citations of 5 articles (5.4 citations/article) | 786 citations of 77 articles (10.21 citations/article) |
| C | 82 citations of 10 articles (8.2 citations/article) | 106 citations of 30 articles (3.53 citations/article) | 29 citations of 8 articles (3.63 citations/article) | 217 citations of 48 articles (4.52 citations/article) |

The table is clear in that the articles classified in B have far more citations per article than the articles classified as A or C. For us as researchers involved in ILSA research, this result is far from surprising. There is a rather

vast amount of research discussing implications of ILSA, but the research more concerned with actually using data from ILSA and discussing methodological questions is more left in the shadow and is not influencing other fields of research in the same manner. Today, when the international scope is conceived as being more important, a normal practice among a lot of research papers is to start a description by saying that international organizations, e.g. the OECD, have an impact on policy and in this using some of the references classified as a B in our prime selection.

The citations among the articles are not equally distributed. A substantial amount of them have few or no citations (PISA have 38 articles without any citations, TIMSS have 11 and CIVED/ICCS have 9 articles without citations). Some of the articles have had a great impact on the scientific community while others are left unnoticed. In a way to show the most influential articles in terms of citations, below we included a table describing the most influential articles. In this it also becomes evident that some of the articles are “crossovers” in the sense that they can be presented both within PISA and TIMSS. It is also evident that the number of citations differs for the same article due to different search dates. This has to do with abstracts being able to contain both PISA and TIMSS and because of our mapping of the field with the help of Discovery they can appear independently in the different steps we made. What are presented below are articles generating more than 20 citations.

Table 27. The most cited articles within respective ILSA search.

| PISA | N. of cit. | TIMSS | N. of cit. | CIVED/ICCS | N. of cit. |
|--|-------------------|--|-------------------|--|-------------------|
| Cross-National Patterns of Gender Differences in Mathematics: A Meta-Analysis. Else-Quest, Nicole M.; Hyde, Janet Shibley; Linn, Marcia C. <i>Psychological Bulletin</i> . | 158 | Cross-National Patterns of Gender Differences in Mathematics: A Meta-Analysis. Else-Quest, Nicole M.; Hyde, Janet Shibley; Linn, Marcia C. <i>Psychological Bulletin</i> . | 154 | Latino Adolescents’ Civic Development in the United States: Research Results from the IEA Civic Education Study. Torney-Purta, Judith; Barber, Carolyn; Wilkenfeld, Britt. <i>Journal of Youth & Adolescence</i> . | 46 |
| Governing by numbers: the PISA “effect” in Europe. Grek, Sotiria. <i>Journal of Education Policy</i> . | 106 | Socioeconomic Status, School Quality, and National Economic Development: A Cross-National Analysis of the "Heyneman-Loxley Effect" on Mathematics and Science Achievement. Baker, David P.; Goesling, Brian; Letendre, Gerald K. <i>Comparative Education Review</i> . | 96 | How Adolescents in 27 Countries Understand, Support, and Practice Human Rights. Torney-Purta, Judith; Wilkenfeld, Britt; Barber, Carolyn. <i>Journal of Social Issues</i> . | 38 |
| The Finnish miracle of PISA: historical and sociological remarks on teaching and teacher education. Simola, Hannu. <i>Comparative Education</i> . | 102 | Efficiency and equity in schools around the world. Hanushek, Eric A.; Luque, Javier A. <i>Economics of Education Review</i> . | 87 | Voice in the Classroom: How an Open Classroom Climate Fosters Political Engagement among Adolescents. Campbell, David E. <i>Political Behavior</i> . | 29 |
| Research on Interest in Science: Theories, methods, and findings. Krapp, Andreas; Prenzel, Manfred. <i>International Journal of</i> | 45 | Student Victimization: National and School System Effects on School Violence in 37 Nations. Akiba, Motoko; LeTendre, Gerald | 67 | Adolescents' Political Socialization in Changing Contexts: An International Study in the Spirit of Nevitt Sanford. Torney-Purta, | 22 |

| PISA | N. of cit. | TIMSS | N. of cit. | CIVED/ICCS | N. of cit. |
|---|------------|--|------------|---|------------|
| <i>Science Education</i> , 50. | | K.; Baker, David P.; Goesling, Brian. <i>American Educational Research Journal</i> . | | Judith. <i>Political Psychology</i> . | |
| Earning Its Place as a Pan-Human Theory: Universality of the Big-Fish-Little-Pond Effect Across 41 Culturally and Economically Diverse Countries. Seaton, Marjorie; Marsh, Herbert W.; Craven, Rhonda G. <i>Journal of Educational Psychology</i> . | 37 | Research on Globalization and Education. Spring, Joel. In: <i>Review of Educational Research</i> . | 62 | International Psychological Research That Matters for Policy and Practice. Torney-Purta, Judith V. <i>American Psychologist</i> . | 20 |
| Differences in Scholastic Achievement of Public, Private Government-Dependent, and Private Independent Schools: A Cross-National Analysis. Dronkers, Jaap; Robert, Peter. <i>Educational Policy</i> . | 37 | Mathematics Teaching in the United States Today (And Tomorrow): Results from the TIMSS 1999 Video Study. Hiebert, James; Stigler, James W.; Jacobs, Jennifer K.; Givvin, Karen Bogard; Garnier, Helen; Smith, Margaret; Hollingsworth, Hilary; Manaster, Alfred; Wearne, Diana; Gallimore, Ronald. <i>Educational Evaluation and Policy Analysis</i> . | 45 | | |
| Educational standards and the changing discourse on education: the reception and consequences of the PISA study in Germany. Ertl, Hubert. <i>Oxford Review of Education</i> . | 37 | Achievement Inequality and the Institutional Structure of Educational Systems: A Comparative Perspective. Van de Werfhorst, Herman G.; Mijs, Jonathan J.B. <i>Annual Review of Sociology</i> . | 41 | | |
| Accounting for the gender gaps in student performance in reading and mathematics: evidence from 31 countries. Marks, Gary N. <i>Oxford Review of Education</i> . | 34 | Patterns of Diagnosed Mathematical Content and Process Skills in TIMSS-R across a Sample of 20 Countries. Tatsuoka, Kikumi K.; Corter, James E.; Tatsuoka, Curtis. <i>American Educational Research Journal</i> . | 40 | | |
| Social segregation in secondary schools: how does England compare with other countries? Jenkins, Stephen P.; Micklewright, John; Schnepf, Sylke V. <i>Oxford Review of Education</i> . | 34 | Institutional and Student Factors and Their Influence on Advanced Mathematics Achievement. Schreiber, James B. <i>The Journal of Educational Research</i> . | 39 | | |
| The politics of international | 30 | The role of plausible values | 39 | | |

| PISA | N. of cit. | TIMSS | N. of cit. | CIVED/ICCS | N. of cit. |
|--|------------|---|------------|------------|------------|
| league tables: PISA in Japan's achievement crisis debate. Takayama, Keita. <i>Comparative Education</i> . | | in large-scale surveys. Wu, Margaret. <i>Studies in Educational Evaluation</i> . | | | |
| School regimes and education equity: some insights based on PISA 2006. Alegre, Miquel Angel; Ferrer, Gerard. <i>British Educational Research Journal</i> . | 28 | Economic Development and the Effects of Family Characteristics on Mathematics Achievement. Schiller, Kathryn S.; Khmelkov, Vladimir T.; Wang, Xiao-Qing. <i>Journal of Marriage & Family</i> . | 30 | | |
| Are between- and within-school differences in student performance largely due to socio-economic background? Evidence from 30 countries. Marks, Gary. <i>Educational Research</i> . | 26 | Differential effects of science attitudes and science achievement in Australia, Cyprus, and the USA. Papanastasiou, Elena C.; Zembylas, Michalinos. <i>International Journal of Science Education</i> . | 29 | | |
| "A Nation at Risk" Crosses the Pacific: Transnational Borrowing of the U.S. Crisis Discourse in the Debate on Education Reform in Japan. Takayama, Keita. <i>Comparative Education Review</i> . | 24 | Students' Progression of Understanding the Matter Concept from Elementary to High School. Liu, Xiufeng; Lesniak, Kathleen M. <i>Science Education</i> . | 29 | | |
| Big Fish in Little Ponds Aspire More: Mediation and Cross-Cultural Generalizability of School-Average Ability Effects on Self-Concept and Career Aspirations in Science. Nagengast, Benjamin; Marsh, Herbert W. <i>Journal of Educational Psychology</i> . | 22 | Gender Differences in Extreme Mathematical Achievement: An International Perspective on Biological and Social Factors. Penner, Andrew M. <i>American Journal of Sociology</i> . | 27 | | |
| Age of selection counts: a cross-country analysis of educational institutions. Horn, Daniel. <i>Educational Research & Evaluation</i> . | 20 | An empirical assessment of the absolute effect of schooling: regression-discontinuity applied to TIMSS-95. Luyten, Hans. <i>Oxford Review of Education</i> . | 24 | | |
| New scalar politics: implications for education policy. Lingard, Bob; Rawolle, Shaun. <i>Comparative Education</i> . | 20 | "A Nation at Risk" Crosses the Pacific: Transnational Borrowing of the U.S. Crisis Discourse in the Debate on Education Reform in Japan. Takayama, Keita. <i>Comparative Education</i> | 23 | | |

| PISA | N. of cit. | TIMSS | N. of cit. | CIVED/ICCS | N. of cit. |
|------|------------|--|------------|------------|------------|
| | | <i>Review.</i> | | | |
| | | Schooling quality in Eastern Europe: Educational production during transition. Ammermüller, Andreas; Heijke, Hans; Wößmann, Ludger. <i>Economics of Education Review.</i> | 22 | | |
| | | National Income, Income Inequality, and the Importance of Schools: A Hierarchical Cross-National Comparison. Chudgar, Amita; Luschei, Thomas F. <i>American Educational Research Journal.</i> | 21 | | |
| | | The paradoxical relationship between student achievement and self-perception: a cross-national analysis based on three waves of TIMSS data. Shen, C.; Tam, H. P. <i>Educational Research & Evaluation.</i> | 21 | | |
| | | Some Characteristics of East Asian Mathematics Classrooms Based on Data from the Timss 1999 Video Study. Leung, Frederick Koon Shing. <i>Educational Studies in Mathematics.</i> | 20 | | |

* Note that some of the articles appear more than once in the different columns. This is maybe not that surprising considering that abstracts can mention both PISA and TIMSS, more surprisingly is the fact that citations might have different numbers. This is due to that the Scopus search is conducted on different dates.

From the presentation above, some observations from before become even more apparent. One has to do with the superiority of some individual researchers. Within referenced research in the CIVED/ICCS sphere it becomes rather obvious that one single researcher is a forerunner, a key figure. Of the five most cited articles one individual researcher has written four of them. It also becomes evident that CIVED/ICCS research has far from the same impact that some of the articles within the PISA and TIMSS sphere have. Among the explicit articles about PISA and TIMSS the same article has the most impact in terms of citations. This is an article describing cross-national patterns of gender differences in mathematics. After that, some differences can be observed. Concerning PISA, several of the most cited articles are articles describing PISA in relation to policy convergence. In this we find many of the key figures of today's comparative education who in many cases have been building both production and careers on criticizing PISA. This is not that evident among the TIMSS articles more concerned with issues more related to the TIMSS content. Because of what we have read, in most of the articles we analyzed, we can conclude with a statement that the articles leading to the most citations are normally articles dealing with overall considerations and because of that are most helpful for citations if you as a researcher need an overall statement concerning e.g. policy convergence, gender differences, differences in attitudes in explaining achievement gaps and so on. As such, ILSA research has been a reference base for

overall statements with an aspiration for legitimation stemming from “*world-wide results.*” As such, ILSA research has become most helpful for the scientific community in legitimating research questions in terms of what can be an appropriate focus from an international view. You can, as a researcher, argue for e.g. equity is a national problem, but if you also state that this is an international problem, you can transform your research interest from an individual or national focus to a much more lucrative focus – suggestions for all humankind.

Who are the researchers?

In our data we have gathered information about affiliation of the researchers writing the articles in our prime selection. We can conclude in that the researchers are normally active in universities as researchers on different levels. There are some independent writers and some attached to international and national organizations, but these are rather few. Among the researchers writing the articles we notice a vigorous bias towards researchers from English-speaking countries. This is may not be that surprising considering the benefits these researchers have in when publishing in English. Another rather striking conclusion is that it is possible to say ILSA research is a rather westernized practice, at least if it is studied in relation to publications in English, peer reviewed journals. Below is a table presenting the affiliations of the researchers in our prime selection.

Table 28. Country of affiliation of the authors of the explicit articles.

| Country | PISA | TIMSS | CIVED/ICCS | In total |
|------------------------|------|-------|------------|----------|
| Australia | 33 | 9 | 3 | 45 |
| Austria | 1 | 0 | 3 | 4 |
| Belgium | 8 | 4 | 5 | 17 |
| Bosnia and Herzegovina | 0 | 1 | 0 | 1 |
| Botswana | 0 | 1 | 0 | 1 |
| Brazil | 2 | 0 | 0 | 2 |
| Bulgaria | 0 | 1 | 0 | 1 |
| Canada | 10 | 7 | 0 | 17 |
| China | 1 | 4 | 0 | 5 |
| Colombia | 1 | 1 | 0 | 2 |
| Cyprus | 2 | 8 | 1 | 11 |
| Czech Republic | 2 | 1 | 0 | 3 |
| Denmark | 1 | 1 | 0 | 2 |
| Estonia | 1 | 0 | 0 | 1 |
| Finland | 9 | 1 | 0 | 10 |
| France | 5 | 1 | 0 | 6 |
| Germany | 16 | 10 | 2 | 28 |
| Greece | 2 | 0 | 0 | 2 |
| Hong Kong | 1 | 3 | 6 | 10 |
| Hungary | 3 | 1 | 0 | 4 |
| Iran | 0 | 1 | 0 | 1 |
| Ireland | 4 | 2 | 1 | 7 |
| Israel | 1 | 7 | 1 | 9 |
| Italy | 3 | 1 | 2 | 6 |
| Japan | 1 | 0 | 0 | 1 |
| Kenya | 0 | 0 | 1 | 1 |
| Mexico | 0 | 1 | 0 | 1 |
| Netherlands | 7 | 16 | 3 | 26 |
| New Zealand | 4 | 1 | 0 | 5 |

| Country | PISA | TIMSS | CIVED/ICCS | In total |
|----------------------|------|-------|------------|----------|
| Norway | 10 | 1 | 3 | 14 |
| Oman | 0 | 1 | 0 | 1 |
| Peru | 0 | 1 | 0 | 1 |
| Poland | 2 | 0 | 0 | 2 |
| Portugal | 2 | 0 | 1 | 3 |
| Romania | 1 | 0 | 0 | 1 |
| Singapore | 1 | 0 | 0 | 1 |
| Spain | 6 | 0 | 0 | 6 |
| Saudi-Arabia | 1 | 2 | 0 | 3 |
| Singapore | 0 | 2 | 0 | 2 |
| Slovenia | 1 | 2 | 0 | 3 |
| South Africa | 1 | 8 | 0 | 9 |
| South Korea | 2 | 3 | 0 | 5 |
| Sweden | 6 | 7 | 0 | 13 |
| Switzerland | 5 | 3 | 1 | 9 |
| Taiwan | 3 | 6 | 0 | 9 |
| Thailand | 1 | 0 | 0 | 1 |
| Turkey | 4 | 2 | 1 | 7 |
| UK | 33 | 11 | 6 | 50 |
| United Arab Emirates | 0 | 2 | 0 | 2 |
| US | 33 | 119 | 20 | 172 |

* Note that some of the articles are doubly booked since they can be registered in more than one Discovery search. That means that some of the articles have the same author and the author might be booked in e.g. both the PISA and TIMSS column.

Within the table the English speaking dominance becomes quite clear — the US especially stands out in importance. Of the 543 total authors, 172 of them are affiliated with the US (equaling 32 percent of all the authors). If you make the same calculation for English-speaking countries (Australia, Canada, Hong Kong, Ireland, New Zealand, South Africa, UK and US) the dominance becomes even more apparent. Of the 543 authors, 315 are affiliated with English-speaking countries (58 percent of all the authors). If you also include European and other westernized countries into the calculation, 497 of the 543 authors are affiliated with these countries (91.5 percent of all the authors). With these calculations as a base, we can conclude that ILSA research is primarily performed by researchers affiliated with westernized countries, at least when dealing with English, peer-reviewed journals.

In which journals is the ILSA research disseminated?

To compensate for the rather narrow systemic review, we have conducted a search consisting of the articles citing our explicit ILSA research. This is a rather time-consuming procedure, but we believe it is time well spent both for widening our description of the research field under scrutiny and also for offering a “fast-track” for other researchers interested in ILSA research. In this we have located 3072 articles making references to our originally 430 articles. Below we present a table of the most used journals within respective ILSA that cite the explicit ILSA research. This is not in all cases the same as the journals with most references to our original articles.

What becomes clear in this table is the fact that the different ILSA to some extent are spreading into different journals even though some of them might be said to have a rather wide scope for discussing different ILSA. One conclusion to be made from the table is that the different ILSA are spreading into different channels and that these channels often favor one of the ILSA. From this aspect it is maybe not correct to talk about a combined field of ILSA research – instead it might be more accurate to talk about separate fields of research

concerning the different ILSA. What stands out is how CIVED/ICCS is spreading, and this is especially evident in the fact that the journal – that is actually a handbook – with the most references to the explicit articles is the Handbook of Research on Civic Engagement in Youth. Once again the superiority might be more visible because that one of the editors for this publication is the specific researcher already mentioned as dominant within the field.

Table 29. The journals where the explicit articles are cited.

| Journal | PISA | TIMS S | CIVED/ ICCS | In total |
|--|-------------|-------------------|------------------------|-----------------|
| Journal of Education Policy | 72 | 5 | 0 | 77 |
| International Journal of Science Education | 52 | 28 | 0 | 80 |
| Comparative Education | 50 | 11 | 3 | 64 |
| European Educational Research Journal | 39 | 2 | 3 | 44 |
| Educational Research and Evaluation | 37 | 40 | 6 | 83 |
| Comparative Education Review | 32 | 16 | 4 | 52 |
| Learning and Individual Differences | 30 | 25 | 0 | 55 |
| Economics of Education Review | 19 | 14 | 0 | 33 |
| Science Education | 19 | 6 | 0 | 25 |
| Research in Comparative and International Education | 18 | 10 | 2 | 30 |
| Scandinavian Journal of Educational Research | 17 | 6 | 0 | 23 |
| International Journal of Educational Development | 9 | 25 | 3 | 37 |
| School Effectiveness and School Improvement | 6 | 20 | 5 | 31 |
| International Journal of Science and Mathematics Education | 11 | 19 | 0 | 30 |
| Journal of Curriculum Studies | 9 | 18 | 5 | 32 |
| American Educational Research Journal | 4 | 16 | 0 | 20 |
| International Perspectives on Education and Society | 11 | 14 | 0 | 25 |
| Handbook of Research on Civic Engagement in Youth | 0 | 0 | 26 | 26 |
| Applied Developmental Science | 1 | 0 | 10 | 11 |
| Education, Citizenship and Social Justice | 2 | 0 | 9 | 11 |
| Oxford Review of Education | 8 | 3 | 7 | 18 |
| American Psychologist | 2 | 0 | 6 | 8 |
| Journal of Adolescence | 1 | 0 | 5 | 6 |
| Citizenship, Social and Economics Education | 0 | 0 | 4 | 4 |
| Educational Research for Policy and Practice | 1 | 4 | 4 | 9 |
| Educational Research Review | 0 | 2 | 4 | 6 |
| Procedia - Social and Behavioral Sciences | 8 | 6 | 4 | 18 |
| Scandinavian Political Studies | 0 | 0 | 4 | 4 |
| Theory and Research in Social Education | 3 | 2 | 4 | 9 |

* Note that the table is based on the most common journals respectively for dissemination of the PISA, TIMSS and CIVED/ICCS articles. This means that the total number of some journals might be higher than the numbers presented in this table. The table is constructed for visualizing similarities and differences, not for presenting the dissemination in total number. This is presented separately in the tables in the appendix.

In conclusion

The different ILSA reports investigated in this systemic review have had a considerable impact on the scientific research if we consider the results published in academic journals. We started out with a rather narrow exclusion and inclusion mechanism, stating that we were only interested in articles making explicit claims in relation to education on a specific ILSA. In this we chose PISA, TIMSS and CIVED/ICCS. In the end we ended up with in total of 430 articles that were classified and further investigated. In order to compensate for this rather narrow scope, we conducted a more thorough bibliometric analysis with the help of primarily

Scopus, but also JSTOR and Discovery. By the end we could locate 3072 articles related to the research field of our interest. We could also show patterns of publications, countries of affiliations and into which journals the explicit ILSA research disseminated.

We have found evidence that the publication of articles follows a pattern that can be expected in this kind of field. The authors of the articles in both the prime selection and in the articles referencing them are primarily university-based researchers with a bias towards researchers based in English-speaking countries and with an even larger bias in favor of westernized countries. The authors are primarily active within the field of education with a bias towards comparative, policy and subjects in schools. Complementary, researchers within the fields of economics, psychology and sociology are active in using ILSA research, but their pattern of publishing in specific journals are more scattered than what is visible in the educational sector. We could also show some problematic issues concerning the research field, meaning that the field is conducted and managed by few researchers with superiority in both publications and opportunities to control the original ILSA. We are not criticizing this, only concluding in that a research field under these premises is to be considered as rather immature and vulnerable.

THE GREYZONE: AN IN-BETWEEN SPACE OF INTERVENTION MODELS IN MCKINSEY AND OECD

Introduction

Part of our review was to capture the relevancing of ILSA research for political decisions and professional work in education. When analyzing this we realized that other actors and publications– outside the field of peer-reviewed research community – had entered the in-between space of research and of the making of educational advices and interventions. Since this was of high relevance for our studies of ILSA translation and contextualization we decided to do a special analysis of what we named a grey-zone between ILSA research and political and professional discourses.

The ILSA grey-zone

In the past decades, new agencies have emerged to interpret the international school system results and make recommendations and specific national proposals for educational improvement. These agencies operate in *grey-zone*, spaces where actors contribute and mediate how the international assessments are to be interpreted and recommendations made about changing educational systems. Operating below the formal radar normally examined when looking at research outputs or policy arenas, the *grey-zone* does not operate by the canons of research communities, nor do they directly bear the same responsibility as policy makers and elected officials. It is what we called earlier; the second face of power in that it organizes principles that serve to delineate the problems of education through its “needs” - statements about educational systems and the targets of its change.

One of these agencies is the McKinsey & Company (from now on referred to as just McKinsey) which publish reports on how educational systems can improve in relation to PISA results and labor market preparation of youth. The other is OECD, the producer of PISA that conducts national policy reviews that draw on its measurement results to make recommendations for national system improvements. If we examine the public discourses about education in Sweden as well as in many other countries, the McKinsey-reports – which were based on PISA-data – were highly visible and played an important role in policy discourses as well as in discussions among teachers’ unions and public debates. The titles of these reports present the promises of comparing the performances of educational systems.

- Barber, M., & Mourshed, M. (2007). How the world's best-performing school systems come out on top. McKinsey & Company.
- Mourshed, M., Chijioke, C., & Barber, M. (2010). How the world's most improved school systems keep getting better. McKinsey.& Company.
- Dominic Barton, Diana Farrell, Mona Mourshed (2013). Education to employment: designing a systems that works McKinsey Center for Government (MCG) www.mckinsey.com/mcg

The OECD has presented reviews on Sweden in earlier documents (e.g. 2011, when Sweden opted to participate in the OECD country reviews with a focus on evaluation and assessment in a decentralized educational system). Here, we however focus on a recent review on Sweden which is to be followed up in *Spring 2015*.²⁶

²⁶ This is a power-point material presented at the website of the Swedish Government. A written report will follow in Spring 2015. Improving schools in Sweden. An OECD perspective. OECD 2015. <http://www.regeringen.se/sb/d/19865/a/258627>

- Beatriz Pont, Graham Donaldson, Richard Elmore & Marco Kools. (2014) The OECD country review, THE OECD-SWEDEN EDUCATION POLICY REVIEW: Main issues and next steps, issued in Stockholm, Sweden, 21 October 2014. Paris:OECD

The OECD Swedish report discussed here was presented to the government to assess the needs of the system and in the subsequent year (2015) to be followed up with a report on how to improve the Swedish schools. Our inclusion of these OECD education policy reviews is as part of the grey-zone and in relation to the McKinsey Reports discussed. Since the THE OECD-SWEDEN EDUCATION POLICY REVIEW was a preliminary report that had no explicit narrative except that the distinctions and magnitudes provided by graphs and charts, our discussion will primarily focus on the McKinsey reports. Where appropriate, the OECD report will be considered. The analysis draws attention to the epistemic congruence of the OECD-SWEDEN and the McKinsey reports through the principles ordering the classifications and comparisons. Furthermore, as the discussion below suggests, there is a strong overlap in and mirroring of epistemic principles between the OECD review report and the McKinsey reports. The identification of the overlap of principles of the *grey-zone actors* is important for systemic review to create visibility.

By means of PISA-results, the McKinsey reports identify what they regard as the best education systems and present reasons for why these systems are so good and what education should do in order to improve or maintain their positions. These reports were often quoted but we did not identify them among our peer-reviewed publications or in the Web of Science. Given this, we concluded that these texts were highly visible on one side (in public discourses of different kinds), but on the other side they were out of the scientific discourses. In a way these publications were “*acting*” in a *grey-zone* between research and policymaking. Thus, we considered it to be of extra interest to do a conceptual analysis of these texts. The purpose of such an analysis is to:

- Portray the system of reason at work in these texts;
- Identify the qualities that made these texts play an important part in the *grey-zone* of policy/science discourses; and
- Discuss different strategies to act in such a grey-zone. Three of these reports are analyzed: we present a short overview of the impact of these publications in the Swedish public discourse (by means of Google search) and in academic discourses (by means of Web of Science and Google Scholar).

The purpose of this chapter is to discuss, first, the contribution of the reports in synthesizing and developing its management models of educational change. Second, it is to think about its principles of knowledge production, classifications, and ordering of interpretations, what we called earlier as focusing on the social epistemology in this *grey-zone area* of reports that serve as mediating schemas between research and policy.

The impact of the McKinsey Reports:

The McKinsey reports have a very varied impact, depending how you measure it. First, on Web of Science they are non-existing, and their authors almost non existing when using this search engine based on scientific publications. Second, according to Google Scholar (2015-31-03) the reports were referred to as follows:

- Barber, M., & Mourshed, M. (2007). 843 citations
- Mourshed, M., Chijioke, C., & Barber, M. (2010). 379 citations
- Dominic Barton, Diana Farrell, Mona Mourshed (2013). 39 citations

However, third, going into the public discourse, as measured by Google the impact is immense, by different kinds of indicators – e.g. 9,5 million hits on “McKinsey and School” over the world.

Looking at the impact in Sweden, a similar pattern occurs, with a large amount of hits - 27 000 – on “McKinsey and skola”.²⁷ Furthermore, the McKinsey reports are having a large impact in the Swedish policy discourse, as *the* reference among political parties and professional organisations.

To our understanding the McKinsey report is outside the scientific publication arena, but had some impact on the scientific discourses as measured by Google Scholar – and here intuitively in three ways: mostly in research fields related to the ambitions of educational change and leadership (e.g. Michael Fullan), to critical analyses concerning the McKinsey reports and the use of PISA in policy discourses (e.g. Frank Coffield). The first is often published in books, the second in journals, and the third in journals and reports.

Instead, McKinsey is having a huge impact on public discourses of different kinds – over the world and in Sweden. This indicates that the grey-zone is of vital importance in education and in much need to be analysed. A substantial amount of citations in these reports use PISA studies as the point of comparison as well as drawing on research publications that follow a similar pattern. A conclusion is that the McKinsey reports have found a space from where to get impact without being scrutinized from a scientific point of view.

Educational change and managing complexity: Contributions.

McKinsey is a global management-consulting firm. The scope of their work moves across private, public and social sectors to address problems that are not only related to specific clients but also to challenges that are considered globally pressing for national and human development. The later societal concerns are not independent of its core function of management but taken as part of the larger corporate public responsibility. In this context, McKinsey contributes to research in areas that address some of the world’s most pressing societal problems. This work has created a nonprofit economic think tank, the McKinsey Global Institute, as well as providing management knowledge for foundations, nonprofits and multilateral institutions on issues related to disease, poverty, climate change, and natural disasters²⁸.

The three educational reports that are examined embody this commitment to provide knowledge that can make a contribution to complex social challenges. The McKinsey reports argue that since school systems do not inquire into how and why they succeed, the expertise of the company can contribute to the international educational improvement through identifying how specific elements of educational systems have broader, universal relevance for providing for “*significant, sustained, and widespread student outcome gains, and examine why what they have done has succeeded where so many others failed*” (see, e.g., Mourshed, Chijioke, & Barber, 2010). An educationalist argues in the forward of this report that the analyses are “*getting at what lies behind the numbers and are thus generating key insights and question, and with this report 'portraying the inner workings of successful pathways of reform given different beginning points'*” [Michael Fuller, np, in Mourshed, Chijioke, & Barber, 2010).

In this respect, the various McKinsey reports examined are important elements to the policy deliberations about the management of educational systems. They provide clear and direct outlines or pathways for policy decision makers to consider how to maintain or improve the organizational features of educational systems. The summaries and interpretations of the vast research measurement programs that link international measures of school performance to the overall management of educational systems provide windows into structures and functions of organizational factors. The reports indicates, for example, what management practices policy makers should pay attention to in terms of the macro organization of educational systems - such as the allocation of financial resources, the level of commitments to change among stakeholders, the need for continuous professional development, and the level of teacher salaries, among others. In the report on education and youth unemployment, the firm’s non-profit organization pays attention to a major international concern

²⁷ Google 2015-03-31

²⁸ http://www.mckinsey.com/about_us

about achievement of schools, issues of youth employment and education that are prominent in relation to current social and economic development discussions.

The mapping of the OECD educational reports and the various McKinsey reports then fills a particular vacuum in the policy field in order to understand complexities through the application of organizational system's models and management theories. The resulting meta-analyses of the existing research begin with the benchmark of OECD's PISA to map different school systems through sets of coherent and clear categories and classifications for thinking about the similarities in interventions.

The discussion that follows considers the principles that order how the knowledge provided in the McKinsey reports orders the complexities and relations of the educational system through its management modeling approaches. At points, we draw attention to the OECD report on Swedish educational policy review to illustrate similarities in the management modeling approach. A more comprehensive analysis of how it relates the presented summaries of Sweden to other countries could not be accomplished until their final report is issued in 2015.

The approach, what we earlier referred to as "styles of reasoning", is in relation to the intent of this systematic review to ask about the results that are presented, their relevance for professional and political decision-making, and principles ordering and classifying the knowledge production and positioning. Thus while the McKinsey reports provide an important element in the mediating of research findings with policy concerns about educational change, the following examines and maps its internal organization of problems, modes of interpretation and solutions of rectification to consider its implications and limits as a lens for addressing educational issues and the problem of educational change.

A Conceptual Approach: Models of systems as fabrications

In this discussion, we focus on the principles of the reason that orders and classifies the McKinsey and OECD reports. These principles are related to its "*systems*" approach and the conceptualization of the problem of comparison and school change. The discussion focuses on the manner of classification to compare schools in national settings, to describe their characteristics and capabilities, and to prescribe the attributes that will produce their change.

To conceptually outline the organizing rules and standards that give intelligibility to the reason of the different reports, we think of inscription of systems' principles through the notion of fabrication. Fabrication, for our purpose, has two different nuances in our analysis. One is a fiction; that is, "*systems*" is a way of inventing particular categories and narratives to order to think about ways to scrutinize the issues and the changes occurring and the challenges that they pose. System is an abstraction that brings together two different epistemological principles for thinking about social life. Schools systems are thought of an organism that grows and changes, a way of thinking borrowed from biology. The idea of systems is also to think of society as a mechanism or machine whose proper alignment of its elements allows for its ordering for administration. The two interacting ideas of systems are analytically in opposition to each other. In the reports under scrutiny, however, there is a practical logic that operates to bring notions of uncertainty (organism) and certainty (machine) together as a mode to think about school performance and administration/interventions.

The second nuance of fabrication is to direct attention to the abstraction of system simultaneously functioning to manufacture or make things. In the analysis below, for example, systems is not longer considered as a way of thinking about schools and change. The abstraction enters into the world as ways of organizing interventions and planning that is to bring the abstraction into actuality.

The focus on "systems" as a fabrication, then, is to engage its double nuances. It is not only a fiction, but manufactures as it enters the world and acts to intervene in people's lives and for people to think of themselves. Systems analysis, to continue with this double nuance of fabrication, is now assumed as describing not only what schools are, but as the principles that order its administration and change that applies not only to institutions but to people. The abstractions that describe school "*system*" in the McKinsey reports and the OECD –Sweden National Report, for example, are treated as determinate systems about what is real and manageable in social life.

The double sense of systems as a fabrication enables us to consider how theories become “actors”. The tendency is to think about people as actors and ideas as descriptions. The notion of fabrication enables us to recognize that the principles that order and classify what is seen and acted can be self-referential and self-authorizing. The overlapping of systems thought and management theory in the reports “act” in shaping and fashioning what are legitimate questions of research, and give boundaries to what is able to constitute remedies that bring the abstractions of the system into existence. We will argue below through the conceptual analysis that this double sense of fabrication is in operation. The systems of schools are taken as material. This double quality entails the translation of complex phenomena into standardized features whose codification is taken as the logically ordered sequences and stages for nations to achieve the efficiency, perfection and equality of the school system.

- There is a further note about the particular use of “systems” as a fabrication in McKinsey and OECD report. In our discussion in mapping explanandum and explanans in ILSA studies, our attention turned to the properties of the research arguments concerning correlations, patterns and regularities over categories and data. The logic or principles that relate internal and external explanans were, we argued, specified through *system as an analytic of thought*. The abstract model of the school as *system* used to capture student performances or achievement gaps gave focus to the administrative characteristics of education, e.g. in terms of elements that can be managed through organizational theories about, for example, tracking and differentiation. The different kinds of explanation combined to order the explanandum.
- The same style of reasoning, we argue below, is embodied in the McKinsey and OECD reports that provide a summary of the research field in order to provide models of change. The systems analysis is made into a problem of management. The problems of research, the categories that drive the analysis of data, the search for equivalences among national “systems,” and the criteria of development and change are self-referential to the theory of management applied as the foundation of the reports. What is given as important are those things that are explicitly manageable and thus can be calculated as equivalences created among different social, cultural and political spaces. This is the use of benchmarks through the reports. *How To Come Out On Top* (Barber & Mourshed, 2007) relates benchmarks as what is construed in the management categories of systems: spending on schools or how teacher quality is defined. The latter means providing rigorous standards and assessment, clear expectations, differentiated support for teachers and students, and sufficient funding, facilities and other core resources.

Systems

The use of notion of *system* described in the reports under review is commonplace in contemporary research concerned with social issues and their administration. Science is given as the problem of administration and, the McKinsey/OECD reports ordering the research and models of change through management theories of social organization. Systems as a general pattern of thought about social interventions received particular importance as a problem-solving tool linked to the programs in the expansion of the post-World War Two welfare states. System “thinking” was used during the War in the mobilization of fighting forces. After the War it was redeployed for the social projects that followed. It tied science to a particular form of problem-solving research. The problem-solving assisted new state initiatives for social and economic reforms. It described and evaluated how new social programs worked, the most effective means for intervention to change institutions and to identify points of weakness for “system” improvement in various social, cultural and economic arenas.

Systems theories were applied as a method of interpretation in the organization of OECD’s approach to study the conditions of education in the late 1950s (see, e.g, B app, 2014; Martens, 2007; Pettersson, 2014); and as a method for interventions in research related to, for example, the *US War on Poverty and the Great Society*, among others. The late 1960s establishment of research and development centers in the US applied forms of systems analysis to the study of school learning, administration, classroom learning and curriculum (Popkewitz, 2011). The language of reform and research embodied salvation narratives of research in service of solving the nation’s ills and providing the pathways for progress -- nationally but also global prosperity. In the US, this was

expressed in the thesis of “*the end of ideology*”; that is, after the defeat of the Nazis, the problem was no longer ideology but finding the paths to a more progressive world society.

The notion of system, however, is not something new in thinking about social relations. Adam Smith theorized about markets in *The Wealth of Nations* (1776), for example, to think about the systemic philosophical claims about universal (human nature) that could be made into an abstraction (the market system). This abstraction could then be given classifications whose elements could be calculated and administered to describe how the abstract system “*worked*” as the fulfillment of the whole. Furthermore, the quantification of the effects or products of these abstractions (labor, national prosperity, rents, and profits) could then become the “*social facts*” that enable comparisons of different “*systems*” (Poovey, 1998).

The numbers that Smith applied; “*embodied [Smith’s] a priori assumptions about what the market system should be*” (Poovey, 1998, p.216). Numbers appeared in Smith’s political economic theory as a strategy to actualize the philosophized fictions of markets as performativity standards instead of descriptions. The sciences would “*solve*” the problem of studying the particulars observed so as to standardize in a manner that could be projected into a resemblance of the future.

The systems analysis that underlies the McKinsey and OECD reports, as with Smith’s markets, is a fabrication whose abstractions could be calculated and administered to describe how the abstract system “*worked*” as the fulfillment of the whole. The classification and “*facts*” of educational success and attainment are shaped and fashioned through the principles of “*the system*” given, a phenomenon that we describe later as the self-referential and self-authorizing quality for organizing the problem, methods, and solutions for rectifying the factors that do not contribute to school “*system*” effectiveness. The quantification of the effects or products of these abstractions are made into and become the “*social facts*” that enable comparisons of different “*systems that embody a priori assumptions about what the school system should be*” (cf. the quote above on how Poovey describe Smiths statement).

Schools as organisms (equilibrium/disequilibrium)

The McKinsey and OECD reports embody the notion of systems research as an experimentalist view of the individuals and society as an organism (see, Easton, 1966; Simon, 1956). Relating earlier biological discourses to social organisms, the principles of systems conceptualized social institutions as organisms with stages of growth or processes of development. The problem of research is to identify and organize the places that did not contribute to the system’s harmony and thus the sites for interventions. This mode of science, as we discussed earlier, is one of “*problem-solving*”. From the organizational theories about “*loose coupling*” to cognitive psychologies about mastery learning, the theoretical integration of systems analysis gave institutional practices their comprehensibility, stability, and functionality (Popkewitz, 2011; Heyck, in press, p.65).

The inscription of systems as an organism is not explicitly stated in the McKinsey reports or the OECD report. It is embedded in the manner of ordering and classifying the problems, methods, modes of analysis and what stands as conclusions/solutions. The school is treated as an organism that can be designed and experimentally intervened in through its organizational management in achieving system’s goals. This notion of organism is evident in the OECD report that examines “*the educational needs*” of Sweden as the starting point of the review (Pont, Donaldson, Elmore & Kools, 2014). The “*needs*” itself is a biological concept, as identified through the comparative data on performances from the years 2003-2012 in reading, mathematics, and science. The data are then correlated to locate Sweden in the series of graphs summarizing the OECD data. Each summary acts as a GPS system (a term that OECD in fact uses) to locate the health or pathologies which limit system performance. The correlations are to identify the needs of the organism in maximizing performance outcomes - such as in measures that differentiate systems in terms of equity with quality. These measures are placed in the report in relation to other elements that are to report levels of disequilibrium - such as teacher and student “*motivation to learn*” and self-responsibility for failure and truancy, educational financial resources in the system, and professional training needs of the teacher. The categories are treated as the system’s organism, each as a site of possible interventions to improve the system’s health (to be discussed in the subsequent April 2015 report).

In the McKinsey reports, the language is of the school as organism that grows and develops. School is an organism likened to the challenge of picking the best organs and best muscle development to get the organism to grow. Systems analysis is to identify the elements of the organization, its points of inefficiencies, and what is needed for ensuring proper growth and development.

This strategy constitutes the underlying theory of school change. In the education to employment report (Barton, Farrell, & Mourshed, 2013), for example, the system elements are described not merely as institutional or organizational. They are essential to the school as "*the material structure of an individual life form*", "*the heart's contribution to the maintenance of the human organism*," a whole with interdependent parts that are likened to a living being. The language is "*the upper strata of the American social organism*" and "*a complex political organism*".

The application of systems analysis is one of equilibrium or about how to hold "*things*" together. The assumption of the system is that it reaches its efficient state when there is equilibrium and harmony in the functioning of its parts. Social phenomena and communication are classified as a continual process (organism) that assumes consensus of the goals and the harmony of the systems. Change is maximizing the utility of its various functional parts. The abstractions of the system are to simplify or reduce the complex phenomena of schools, culture, and political regimes (the latter is an explicit category of McKinsey) in order to make them manageable for research and social interventions.

The McKinsey research codifies and standardizes to identify points of intervention. Interventions assume the validity of the parts and the mapping of their functions. For example, to identify school systems that have greatly improved, bar graphs are presented to identify criteria of school systems to assert the developmental logic. The logic is one of scaling, creating a continuum of value about "sustained improvers (1983-2007; promising starts, 1990-2007 (years very according to country/city); type of system, such as nation, province, district, network, wealth, performance level, size of system (# of schools)" (Mourshed, Chikioke, Barber, 2010, p.14). Performance Stages are identified with sets of discrete markers about development - such as rating from poor to fair, good and excellent. When moving from great to excellent, less guidance is necessary. The latter entails, for example, having looser guidelines, with peer-led creativity and innovation that entail discrete management actions that are related to performance - such as revising curriculum and standards, ensuring appropriate rewards and remunerations, and building technical skills of teachers and principals.

Methods stipulate the factors or elements that interact to make up the system and then calculate and measure to enable social interventions that actualize the desired elements and achieve the harmony of the system. The coding and standardizing of the elements form equivalences from which to calculate differences according to the standards given in the equivalency measures. Stages of performance spectra are invented on "*a universal scale of calibration*" to normalize, for example, several different international assessment scales of student outcomes discussed in education literature (Mourshed, Chikioke, Barber, 2010, p.7). The universal scales symbolize the development and stages that enable the growth of the organism. The models of schooling, for example, link measures of achievement, teacher training, learning and personality theories, school organization, and parent participation to an organism whose management creates a more effective and efficient learning organism.

The assumption of equilibrium entails its opposite, disequilibrium. The stages of development are stipulated as a continuum of value. This stipulation is expressed in the very titles of the reports: *How the world's best-performing schools systems come out on top and; How the world's most improved school systems keep getting better*. Successful schools are benchmarked, a word that assumes a clear standard of performance from which to measure difference. The benchmarks serve as the ideal to construct a scale of difference. Research identifies the points that produce the pathologies that disrupt the harmony of the desired models of the system. These pathologies create the disequilibrium that is to be eliminated and the stasis or balance to be achieved, i.e., utility maximization in relation to its desired outcomes. This disequilibrium/equilibrium gives exemplification of the *achievement gap* among nations.

Discussions of PISA in the McKinsey reports are directed to instruct school systems to correct those elements that contribute to its disequilibrium. The delineating of steps to achieve success or to bring the consensus between educational practitioners and employers to reduce youth unemployment entails the identification of pathologies that produce dysfunctions, such as the lack of coordination between educator and

employer. Each general element, category or segment described as essential for youth employment, for example, is researched to identify finer distinctions that serve in the hierarchy of stages and development in order to bring equilibrium between education and youth employment. The stages of development are not only organizational factors but are lined with psychological qualities of youth that normalize what is functional and dysfunctional for employability – such as disengaged, disheartened, well-positioned or too poor to study (Barton Farrell, & Mourshed, 2013, pp.32-33). In the OECD report, the graphs of performance serve as GPS to locate Sweden's health (normalcy) and pathologies in the particular categories taken to represent the functioning of educational systems (Pont, Donaldson, Elmore & Kools, 2014).

As with Adam Smith's fabrication of "*markets*" as a system, the given classifications of the school system are standardized, calculated and measured to define the "*social facts*" that enable comparisons of different "*systems*". The classifications and measures are to permit its projection into a resemblance of what fulfills the future school (and teacher and child). The classifications connect different and discrete things for comparison of the internal workings of institutions and the whole along with the relation of the internal distinctions to external patterns, such as family influence on children's achievement or the relation of education to employment. The assumptions of change are, paradoxically, bound to stabilizing and maximizing the qualities found in existing classifications and ordering principles. Its focus on stability is part of the assumption of equilibrium/disequilibrium. That assumption when tied to the idea of organism and "*needs*" creates a hierarchy of values. These values embody distinctions between pathologies which hold back the health of the system and the normalcy given as the abstraction of the system of management. These pathologies are not named as such but are embodied in the GPS system that locates not only the pathways for organizational management but also the social and psychological characteristics of the children and families who succeed and fail (in such distinctions as motivation and self-responsibility).

Self-referential/Self-authentication

The research under examination is self-referential/self-authenticating. Self-referential in that the reports embody particular a set of principles that order and create boundaries through which the problems, recognizing and distinguishing the objects of schools and modes of finding out are made understandable and manageable. The self-authenticating is to how the abstractions of the "*system*" frame not only what is looked at but loop back to determine what are valid "*solutions*" and which endorse the components of the system as the only legitimate way to think about what exists and its change. In the case of McKinsey, the self-authenticating is the making of the problem of schooling as inadequate management of the systems that produces disequilibrium. The reading of the OECD-SWEDEN report is through how the images presented in the charts and graphs embody narratives about system characteristics and which have a similarity to management categories that form the McKinsey theory of change.

The self-referential/self-authentication is an inherent quality of science itself, what can be thought of as the paradigmatic qualities that organize research. Yet, there are differences. Science is built on refutations if we use Karl Popper's famous book; whereas, the McKinsey reports are organized for certainty and as a self-confirming system. This certainty, as we argue below, has four qualities that give the reports a particularity that is outside of the norms of science. One, there is no room for questions outside of the principles of ordering and classifications given. Two, it is assumed the abstract models of the school system are mirrors of reality. Three, they claim the knowledge to achieve system performance goals is already at hand. The problem is to use that knowledge to identify the paths or "*highways*" to achieve its equilibrium. And four, the only problem of their research is to get system actors to manage more effectively the predefined paths on highways and get to the destination's endpoint.

These four qualities are themselves embedded in the self-referential/self-authenticating characteristics of McKinsey through its principles that organize "*systems*".

The categories and classifications inscribe finding the right management procedures as related to seeking the harmony of the organismic system. The elements and solutions are enclosed within and actualize the abstract models in order to achieve better managing of the system. When class size is found as not influencing school achievement, better management procedures are deemed as the solution through, for example, hiring the right

teacher. The right teacher is defined through the management procedures of setting standards and using assessment tools (Barton, Farrel, & Mourshed, 2014). Or, the management category of “*leadership*” is designated as “*most important in igniting success*” (Mourshed, Chijioke & Barber, 2010, p.140).

Differences are defined through the prior criteria of management and its scaling and developmental stages that are metaphorically and literally reported as the path of the given “*highway*” to improved management techniques. The research task is to find proper and efficient routes to maximized performance. Research provides “*the element specific to individual systems and those of broader or universal relevance to understand why some systems have registered significant, sustained, and widespread student outcome gains, and examine why what they have done as succeeded where so many others failed*” (Mourshed, Chijioke & Barber, 2010).

The reports function as a rubric to establish system harmony. The research is a continually two-step process of data collection: assembling international data with international assessments of nations as the benchmarks and then identifying stakeholders for interviews that confirm the categories of benchmarks as the origin of the problem. The statistical categories are used as the basis of interviews with system actors. The interviews are the self-authorizing devices to develop scales within each category of management that circulates as the theoretical presupposition in each report. The classifications are to identify the scales of development to achieve maximum system management, that is, to produce equilibrium of elements to produce system goals. The categories of the statistical analyses become the hierarchy of implementation that classify the interpretation of the interviews, for example, with 200 system leaders and visits to view all 20 systems (Mourshed, Chijioke, & Barber, 2010) or employment initiatives from 25 countries, then a survey of youth, educational providers and employers in nine countries (Barton, Farrell, & Mourshed, 2013).

The highway metaphor is used in the 2013 report to think about the problem of disequilibrium: “*to get rid of potholes, make educators and employers part of the solution by providing ‘signs’ and concentrate on patch of pavement ahead*” (Barton, Farrell, & Mourshed, 2013, p.54). The problem is to reach destination that is already known in a reasonably efficient manner. The highway is not only a metaphor but embodies the double sense of fabrication. It is the path to achieve the system’s equilibrium. The McKinsey’s language of highways, for example, posits them as the correct paths to “*scale up to delivery sustainable skills*” (Barton, Farrell, & Mourshed, 2013). The highway enables safety, efficiency, and health of society.

The research calculates and validates the categories given in the abstract and symbolic model of the system. The interviews are tied to the categories already existing in the management theory. In each report, spectrums of performance are developed to organize processes that serve solutions, given the category “*excellence*”. The assumption of *How the world’s most improved school systems keep getting better* (Mourshed, Chijioke, & Barber, 2010), for example, identifies the continuum of practices that McKinsey argues produces successful achievement. This entails, for example, identifying the national school systems with significant sustained and widespread results in student outcomes from 1980 onwards in developed and developing countries and classifying them on a grid, such as *sustained improvers* (e.g., Singapore, Ontario, Poland) and *promising starts* (e.g., Madhya Pradesh, Minas Gerais, Western Cape), as substantiated by interviews and observations (Mourshed, Chijioke & Barton, 2014).

The spectrum rests on “*a universal calibration*” developed by “*normalizing several different international assessment scales of student outcome*”. Improvement is to establish collaborative practices, develop mediating layers between schools and center, and architect tomorrow’s leadership.

The continuums are expressed as highways that, if followed, will bring nations to the top of the continuum. The scaling, continuums of development, and system assumptions of consensus and harmony give certainty to what are described as solutions. Educational attainment, it is asserted, can be achieved within as little as six years (Mourshed, Chijioke & Barber, 2010, p.14). In the study of youth and employment, this certainty entails the underlying assumption of the specifications of the school system and “*its outer environment*” as a rational system where humans pursue their rational interests. The rational choice principles cross the three reports in its expression of the school system: how to make the right choice so as to prepare for entrance on the highways and to move to high achievement (Barton, Farrel, & Mourshed, 2013). The given certainty of the universal model or ideal type makes contexts as the stages that lead to fidelity and harmony in the functioning of the systems (Barton, Farrel, & Mourshed, 2014).

The highway is what enables them to pursue those interests that are in fact the interests superimposed on the analytics through the abstraction given as the “*system*”. For example, the study of labor and youth assumes that one element of the system is its communication elements. When it is found that youth do not use the information available to them, the solution is to give youth better and more efficient information (Barton, Farrel, & Mourshed, 2014). The report identifies ten areas of impact that can provide system efficiency (such as professional development, accountability, and learning models), which is then disaggregated into 60 unique subareas, such as the management category of accountability. Analytical distinctions are created as the points for interventions to eliminate the disequilibrium through following “*the highway*”. The strategies are posited as changing existing structures, resources, or processes, and also as identifying which agents the interventions are to act upon - such as the principal, teacher, or student (Mourshed, Chijioke, & Barber, 2010).

The analyses are to provide finer distinctions for the prior standardizing and coding procedures. The solutions are predefined: better assessments, from central to empowering teachers and local control, lifting teacher status, collaborative practice, “*mediating*” centers, and defining the architect of leadership to maintain continuity. There are no surprises except as they relate to the system coordination.

Developing equivalents as universals to actualize

We will focus on the McKinsey reports. The comparative projects in the McKinsey and OECD reports entail overlapping strategies to make otherwise incommensurate school contexts as commensurate through developing categories of equivalence in the classification of “*systems*”. Two strategies evident in the reports are discussed here. The first is to develop classifications that serve as universals from which to measure differences. The second is to define context as variations that contribute or disturb the system equilibrium.

The reports’ research problem is to create sets of equivalences so as to enable comparisons from different databases. Using the abstract model as an ideal type, distinctions are calculated and measured as administrative units. The units are then placed into a continuum of values that is to actualize the system’s harmony and equilibrium. Statistically, the reports develop new algorithms for comparing differences. Different and independent assessment scales are placed into single universal scales that normalize and compare, such as comparing educational spending per student (Barber & Mourshed, 2007).

The scales and “*highways*” deploy the abstraction as the ideal type, described as “*best practices*”, from which to consider difference within the schema of management. Agents are described as those categories of people in the system that require the implementation of more efficient management procedures; that is, to address those parts that are out of sync now with the optimal vision of the model. The abstraction of the ideal model of school defines comparative sets of values in order to understand difference.

The comparativeness requires, however, creating analytic devices to normalize differences by statistical secondary order measures that take different data sets and put them together as equivalencies. Difference is the calculated distance from the universal that serves as its norm. The analysis requires homogenizing the differences by developing stages of a performance spectrum on “*a universal scale of calibration*” (Barton, Farrell & Mourshed, 2013, p.7). The report about how schools keep getting better as well discusses the need to develop a higher order statistical measure that can normalize several different international assessment scales of student outcomes discussed in education literature (Mourshed, Chijioke, & Barber, 2010).

Comparison is about differences reduced to sets of equivalences that enable the comparisons. Different and independent assessment scales are rescaled, for example, into a single universal scale normalized and compared from the universal scale that have similar educational spending (USD 6000 per student) (Barber & Mourshed, 2007).

The standardizing and codifying of the distinctions erase difference through developing another statistical level from available data sets in order for McKinsey’s mapping of the system to function. The map’s routes are formed through establishing variations as a continuum of values directed toward achieving the universal characteristics of the school system. Variations, however, are standardized and thus similar sets of interventions are shaped and fashioned through the management categories. Along with talk of “*tightness of central guidance to schools*”, is recognizing flexibility and loose guidelines needed for “*highly skilled educators*”. The

flexibility and guidelines, however, are a form of goal steering in the system where freedom and flexibility is interned and enclosed through the principles of the “*organism*” seeking consensus and harmony.

This is evident when the scaling and norming procedures are examined as defining the spaces of actions. The scales to differentiate are expressed as a continuum of value in achieving system harmony. The scales rate elements on a continuum from “*poor/fair to good*”, “*good to great*” and from “*great to excellent*”. In the report on how school systems are improving, the scale is given as a clear and linear progression internal to each category and across categories (Barton, Farrel, & Mourshed, 2014), such as:

- *Fair to good*: consolidating system foundations, high quality performance data, teacher and school accountability, appropriate financing, organization structure, pedagogical models;
- *Good to great*: teaching and school leadership as a full-fledge profession, necessary practice and career paths as in medicine and law; and
- *Great to excellent*: more locus of improvement from center to school, peer-based learning, support of system-sponsored innovation and experimentation.

It is within this space of actions to apply system imperatives, which is this discussion of looser guidelines and “*peer led creativity and innovation*”. Peer-led creativity and innovation inside schools are management tools that become “*the core driver for raising performance level*” (Barton, Farrel, & Mourshed, 2013, p.20). The governing at a distance seems varied and diverse but is embodied in the “*system*”.

Context in the reports is situated within these strategies of finding equivalence and variation. Reference to context is what contributes to the dis/equilibrium of the management system. At one level, the reports continually reference diversity, such as studying “*employment initiatives from 25 countries, selected on the basis of their innovation and effectiveness*” and engaging in “*a survey of youth, educational providers and employers in nine countries that are diverse in geography and social economic context*” (Barton, Farrell, & Mourshed, 2013). The diversity and innovations are reclassified as contextual in the sense of how to apply the universal “*to navigate the challenges in their context and to use their context to their advantage*” (Mourshed, Chijioke & Barber, 2010, p.20). Context is a tactic the system-leaders use in tailoring the set of the interventions needed on their performance journey to their specific context. The aim of system learners in contextualizing is said “*to gain requisite support of the various stakeholders*” for the interventions being made.

It is important to note that those explicit classifications that are manageable and thus can be treated as equivalences homogenize different social, cultural and political spaces. Context embodies ahistorical and transcendent qualities tied to the abstraction model. Benchmarks in *How to come out on top* (Barber & Mourshed, 2007) are management categories of systems that can be given statistical calculations as standards and expectations – such as teacher assessment tools, differentiated support for teachers and students, sufficient funding, and facilities. The management characteristics are calculable and measureable so as to override considerations of culture, politics, context and the foreseen in the governance of the system. While the latter qualities cannot be easily put into standardized and codified categories, the analysis creates superordinate algorithms in which to create equivalences that neutralize the indeterminate qualities of social life (Barber & Mourshed, 2007, p.13).

Context comprises the various elements of the categories defined as external to the school system but contributing to its utility. The categories of context are parsed from the standardizing procedure. The thinking of context in this manner is methodologically justified in the reports as the need to standardize in order to “*exactly*” define “*what one system can learn from another*”.

The reduction to the existing categories that define the system inscribes the problem and solution of school success in a self-referential and self-actualizing mode of thought. The universal “*benchmarks*” and the establishment of statistical procedures of equivalences drive what are context-specific but whose variations are inscribed in the scales and categories posited as what makes the system.

The search for certainty

The reports continually suggest certainty through its notions of highways and distinctions that make the problem of change as that of management. The “*play of language*” in McKinsey and the OECD reports are only reporting findings from their summarized research (abridged in graphs) and interviews for the McKinsey report. McKinsey asserts, for example, it is merely mapping the “*systems*” because schools have not done it (Barton, Farrel, & Mourshed, 2014). While there is a language about “*suggesting*” or merely reporting, the organization and purposes of the reports make declarative statements that McKinsey has identified “*the best practices for achieving the implementation of the management schema irrespective of the culture in which they are applied*” (see, e.g., Barber & Mourshed, 2007, np). The purpose of the research asserted in one report is to examine 20 systems “*in action*” so as “*to give precision to clusters of intervention that are not merely the ‘results phenomenon’*” as was PISA. The certainty is also expressed in such statements as the reports are; “*getting at what lies behind the numbers and are thus generating key insights and questions, and with this report portraying the inner workings of successful pathways of reform given different beginning points*” (Fuller, Preface in Mourshad, Chijioe & Barber, 2010, np).

The system is presented as hyper-rational and taxonomical. There is presupposed a defined and known logic to the system’s reform. These are given as the pre-defined stages, intervention levels, and developmental sequences that are postulated as leading to effectiveness. The schemas of interventions are scaled as hierarchies. They embody general system categories and internally to the specific categories within the given levels of development. For example, there are precisely six interventions of revising curriculum and standards, ensuring appropriate reward and remunerations structures for teachers and principals, building technical skills of teachers and principals, assessing students, establishing data systems, and facilitating improvement through the introduction of policy documents and education laws (Mourshad, Chijioe & Barber, 2010).

The overriding principle is that the knowledge of how to achieve and to change is already known and the only problem left is to find the correct pathways. This is quickly asserted in the Executive Summary about the top performing countries (Barber & Mourshed, 2007, np): “*The experiences of these top school systems suggest that three things matter most: 1) getting the right people to become teachers, 2) developing them into effective instructors and 3) ensuring that the system is able to deliver the best possible instruction for every child*”. It is assumed that McKinsey has identified how schools systems can improve, and the only issue is whether schools systems can actualize what is already calculated and measured.

The benchmarks are given as the proved ideal type to compare and identify where nations are on the stages of development or how the system can be brought to actualize the ideal type, such the relation of education and youth employability. The GPS system is there for individual school systems to develop their navigational strategies to be the best schools systems that come out on top. The reports compare and recognize difference in school systems on a continuum to understand, for example, “*the route by which others can get there*” (see, e.g., Barber & Mourshad, 2007, Preface, np). Intervention patterns are strikingly consistent across nations pursuing similar outcomes. The relation of youth employment and schooling is stated as lacking adequate data initially in the report. But in the reporting and calculating of the available sets of data, a different set of conclusions emerges. The great variations are reordered into categories that universalize to present strategies of intervention that will solve the problem of youth unemployment: “*What precise interventions occurred in each school system and when, and how these interventions interacted with each other and with the system’s broader context to deliver better outcomes for students*”. The questions asked assume that the knowledge already exists for solving the youth unemployment problem.

Illusion of democracy and equity

The language of universal stages, processes and procedures portrays the knowledge at hand so all nations can reach the top. The problem is for stakeholders to follow the given highway. One can read the McKinsey descriptions of highways and their prescribed paths as democratic. Democratic in that if all nations and actors follow the given highways and install the proper planning and coordination, it is possible for every school system and nation to achieve equality among schools and nations. The image and narratives entail single,

universal scales for development and what is needed is to follow the standards given, such as the use of assessments on teachers (Mourshed, Chijioke, & Barber, M., 2010 p.14). The universalism of the claim is so broad and abstract to project that the entire problem of nations, schools, and economy for functioning well and productively is the proper implementation of the stages and “highways” through the schemes of management.

These universalistic claims about what must be done are accompanied with claims that there will be no differences and differentiation if the schemas are followed. The “main message,” for example, in the report on how schools are improving is that any school system can improve from any starting point, and to do so, system leaders must integrate three aspects when developing and implementing improvement: “(1) to assess the status quo or performance stages that identify where schools system can start and go; (2) design intervention clusters necessary to reach higher performance stages; and (3) the system’s adaptation of intervention cluster that takes into account history, culture, politics, and structure of school system and nation” (Mourshed, Chijioke, & Barber, M., 2010, p.18).

Exclusion is simply managing the system better for inclusion. For example, the McKinsey report on employment, youth and education asks: (1) how can providers, governments, employers encourage students to take and finish the right courses to prepare them for where the jobs are and (2) how can social perceptions about vocational training be addressed? The solution is to develop detailed and comprehensive information to “struggling youth” (Barton, Farrel, & Mourshed, 2014, p.60).

It seems so straightforward and obtainable for all. Yet, all the data that is available about school reforms and change would suggest that the logic of change and development is a social, cultural and political process that has no guarantees. If we maintain the language of systems, the logic of practice is much more complicated than the practice of the logic of management. The commonsense approaches of the reports are illusions about abstractions that, however, can loop back into everyday life as “acts” that shape and fashion what is taken as real.

Inclusion/Exclusion

The construction of equivalences for comparison contains a different layer of organizing difference. This entails the inscription of consensual norms about what constitutes the successful school and child. As we discussed earlier, there was a continual scaling and hierarchy of values that described the development of the school according to the desired model of the school system. Included in these principles are norms about the child who does and should succeed if the school actualizes this model.

While the reports are presented about nations and their relative ranking in international measures, this ranking entail categories and classification about the child, family and community derived from the assumed consensus of the system’s models. The field of scales and management values that define the system expresses references to the kinds of people – children, family, and community – who are the target of the developmental scores of the nation. These references assume a consensus about what is measured and what needs to be done by actors in order to change the school actors.

The assumed consensus entails double gestures or unarticulated ideas or sentiments that express the hope of education in finding solutions that integrate the individual into society and allows for its general development and progress. The double gestures silently travel in the categories and distinctions of the reports. The double gestures are instantiated in the principles of disequilibrium/equilibrium given to problem, methods and solutions that classify the characteristics of children and teachers. The double gestures are instantiated with the PISA measurements, which disproportionately identify high achieving countries/cities that have strong homogeneous populations – economically, religiously, and socially/ethnically – Shanghai, Hong Kong, South Korea, Singapore, and Finland, the latter with the smallest immigration populations within the Nordic Countries. The distinctions with Minerale General or Boston among difference immigration populations are homogenized through the statistical measures of identifying equivalences while ignoring how the universal criteria of achievement (in PISA’s case what is considered practical knowledge for the future) embodies internal differentiations among populations even when there are efforts to control the factors.

Embodied in the modes of reasoning are the hopes of realizing the model by changing children, parents and communities so they can “improve,” achieve higher or for youth employment. Teachers are the agent of

change. Accompanying the hopes are fears about the dangers and dangerous populations that threaten full utility of the system. The double gestures operate to differentiate and divide populations within nations. This is tacitly present in the scaling and hierarchies that project the lower categories in the taxonomies as threatening the development of school but also as a society that is given as universal through the model applied.

The reduction of complexities to those of rational management “*systems*” makes it seem that “*all*” national systems can be equal through the application of categories that recognize difference and this inscribes differences so that all can never be at the “*top*”. The characteristics of people who succeed and don't succeed are inscriptions of particular distinctions and divisions. The qualities form a continuum of value about the teacher and child hoped for through the models and what are feared as dangerous to the system's functioning.

The double gestures are evident when issues of poverty are discussed as a problem of management procedures. The reports discuss the problems of poverty by classifying the poor and racial groups as lacking particular psychological and sociological skills for successful participation. The recognition of difference to include defines those populations as difference and thus not of the average. This is evident when the subjects of the research are the trilogy of urban populations: the child, family, and community. These distinctions appear, for example, in the linking of organizational development stages with the psychological qualities of youth that normalize what is functional and dysfunctional for employability - such as disengaged, disheartened, well-positioned or too poor to study (Barton Farrell, & Mourshed, 2013, pp.32-33).

The social and psychological categories involve double gestures. The hope is for the child to be motivated and to have parents who have certain interactions and communications with the child; the language of “*lacking*” signifies not only the system's disequilibrium to be corrected. It also embodies norms that are outside of the assumed consensus that shapes and fashions the distinctions of the “*system*” itself. The fear is embedded in those who do not have the norms for system participation - such as the poor as lacking political efficacy, socially disadvantaged and at-risk.

In the youth study, for example, the report begins with a discussion of higher education and includes all populations from universities through vocational education. The purpose of the labor/education report is stated as “*Around the world, governments and business face a conundrum: high levels of youth unemployment and a shortage of job seekers with critical skill, there is a gesture to speaking about 'all' youth and employment*”. But that is not the case. The percentage of youth across the nations studied is given as a universal category from which differences are established. The general analysis quickly focuses on the poor and other categories of populations who go to vocational education as the center of analysis and recommendations. The principles of system functions, structure, and developmental processes become the reference for thinking about the teacher, but also includes is the trilogy as the subject of research and what are “*robust*” data in current teacher education research concerned with practice. The norms and its distinctions of difference universalize the particular group that is unemployed as all youth and then focus on those who are not college graduates, treated as dangerous populations on the highway and how to make the dangerous kinds of people (the unemployed youth) as no longer a threat.

Cherry picking

There is a “*cherry picking*” of data to argue what works. In part, the selection of what constitutes “*data*” for comparisons can be understood in relation to our prior discussion of the self-referential and self-authorizing qualities of the systems principles that organize the reports. But there are also internal omissions. For example, the construction of teacher stipend of \$50 for professional development as a best practice uses the reports in Finland and Singapore but others that do not use this are not reported. When talking about sustaining, it is stated that it is not merely explicit structure and management approaches that matter but also to how teachers think about teaching. The reference is to the educationalist Lee Shulman's work about moral structure or dimension of beliefs that have no correlation to the argument at hand. To look at styles of difference in sustaining successful school systems, the gamut of economic and political regimes and geographical places are homogenized as the same - such the nation of India described as regimented and the province of Western Cape as having flexibility in their district. In the labor report, the criteria of statistical relevance shifts. At one point, discussion is about how only half of youth believe post-secondary studies improve employment. At different

points in the discussion are lower criteria of significance (Barton, Farrell & Mourshed, 2013). And as discussed above, the labor employment report discussed the issue initially as post-secondary education only to then consider the management of vocational training as the “*problem*” (see. e.g. Barton, Farrell & Mourshed, 2013, p.61).

Conclusions

Our analysis of the grey-zone considered the in-between actors in the space between research and policy. The focus was on the particular sets of principles about educational phenomena and change that circulated in the grey-zone reports of McKinsey and OECD. The examination of the grey zone was on the style of reasoning. The chapter explored the principles to organize what is known but how one should act and think of solutions as important to contemporary issues of education. The approach sought to understand in inscription of research of ILSA as its forms of codification enter into the models of policy planning and educational intervention.

But as noted, the principles of reasoning were not only the grey-zone. They were linked to the assumptions given in the research data discussed in prior chapters. As discussed earlier, the research programs of OECD and TIMMS, for example, speak of themselves as calculating and measuring certain aspects of the performances of students in schools, the measurements are linked beyond the comparative statistics to elements such as the social and psychological qualities of the child who succeeds and fails; and the qualities of the teacher that fosters or hinders achievement. The sets of relations that are described in the research and the McKinsey/the OECD-SWEDEN reports discussed are organized around the double sense of the fabrication of schools as a system. The abstraction the school as a system not only for describing and collecting data about school performance and its assessment but also formed models of standardizing and planning change.

These principles related to system entailed the presumption of consensus about the purposes and the knowledge being codified. The principles of systems were embodied in the methods of comparison functioned to translate the complex phenomena of school practices into standardized features whose codification to logically order the sequences and stages for nations to achieve the efficiency, perfection and equality.

This assumption of consensus made it possible to think of the system characteristics as indicators of what is useful and relevant. When criteria of useful and relevant knowledge are spoken, they are bound to self-referential/self-authorizing reasoning of the system. There were overlapping principles of “systems” that shaped and fashioned the legitimate questions of research, what constitutes data, and form boundaries to what are to constitute the norms of planning and interventions that remedy the problem of school performance. It is within the boundaries of the abstractions that values of use, practical and relevant are produced.

Relevance was linked to the principles of the school system as continually striving for a state of equilibrium. The elements of the system that disturbed the consensus and harmony, it was argued, presented themselves as the system’s disequilibrium. Such elements were the lack of adequate professional selection, assessment and development procedures.

The identification of disequilibrium established norms and distinctions to constitute the system’s pathologies that required interventions and remediation. The pathologies were given as the system’s context or environment. Context was the site where effectiveness and efficiency interfered with the school system performance. The categories of context were those that could be administered as influence system outputs. School contexts/environments as the intersection of historical, cultural and political qualities that created ambiguities and uncertainty were removed. In their place give schools were distinctions that seemed independent of particular national context but which could be applied to differentiate and compare, such as benchmarking standards, teacher starting salaries related to the OECD average or GDP per capita, norms about comparative student outcomes.

Finally, the comparative reasoning in the research and intervention models were considered as double gestures. One layer of the reports was the articulation of the hope of equality both at the organization level of nations and in the social level. At the national level, the development of universal categories from which to compare different national school systems gave the statistical measures an implied certainty for the development of intervention models. It also brought values of democracy into the comparisons through the models of intervention. The assumption was that all nations were equal and could rise to the top if they

followed the recommendations and “highways” proposed. At the social level, the universal categories established social and psychological distinctions related to the child, family, and community that differentiated success and failure in school outcome measures. These “background” measures worked within the system’s principles of consensus and equilibrium that embodied the points of disequilibrium and pathology. It was argued that the recognition of comparing from the universal norms and distinctions provided differences and divisions. The divisions were pathologies of populations dangerous to the system’s models and highways and feared if not changed.

CONCLUSIONS AND DISCUSSION

Introduction

This is a report from a systematic research review on international comparisons in education by means of International Large Scale Assessments (ILSA) such as PISA and TIMSS. We were asked by the Swedish Research Council to carry out such a review as part of the implementation of the new Swedish Institute for Educational Research. Given this we developed a model for doing such a review and analyzed how the review functioned. To the institute a mission is to organize and present knowledge of relevance for different stakeholders, such as education professionals and policy-makers. Thus we also analyzed inquiries of translations into practical reason and contextualization of ILSA research.

Reviewing the reviews

A model was designed as a combination of aggregative and configurative characteristics (Weiss, 1979; Gough, 2012) in the making of systematic and transparent research reviews. We started with an identification of a large set of research publication in the ILSA field – more than 11 000 texts were identified by means of search engines. Of these we decided to choose the PISA and TIMSS research programs plus the CIVED/ICCS research program on civics education – in sum more than 8 000 publications. This was filtered in a series of steps in accordance with our design.

Given the task to assess research quality we only included peer reviewed scientific articles. Using this criterion excluded almost half the number of publications. Thus, the demand of scientific quality assurance did matter much, which we consider as highly relevant in a research review, but we would later have tested implications of revising this criterion given a less strict time limit. For instance, using this scientifically focused criterion we did not identify “grey-zone texts” that later turned up to be highly influential in policy and professional discourses but not visible in the first steps due to our search design.

When using another criterion – stating that we should include only primary research doing international comparisons – an additional large share was excluded. We did not see this as a problem, since we actually focused on international comparisons, but it showed the importance of transparent criteria when doing systematic research reviews.

Furthermore, by the use of the searching devices publications in Swedish were almost non-existent. Observing this we did additional search of publications by means of Swedish research engines (e.g. Swepub) with some improvement in number of publications from Sweden and in Swedish language. Given that we know about high quality ILSA research, this points to a visibility problem for Swedish research in such reviewing that we carried out. This is in need for further analyses – e.g. in terms of communication strategies.

Review results – arguments and conclusions

Important in mapping and synthesizing research was to capture arguments and conclusion in a broad field that varied in terms of study objects as well as knowledge objects. We turned this into matters of explanandum and explanans (von Wright, 1983, 2004). Given this we could map and synthesize relevant research publications based on the arguments presented in the research.

A broad result concerns what to be explained – i.e. the explananda. We noted in the reviewed articles a very large share of identifications of achievement gaps over population taxonomies – e.g. classifications in terms of social class or gender – pointing to inequities and how these inequities were associated with different kinds of education measures or to contextual variations – e.g. gender inequity coefficients in different countries. To a much lesser extent differences in efficiency were analyzed – e.g. school performances over educational measures or school systems. We also found research analyzing how to redirect or govern students into certain careers – often in science – being highlighted in some studies.

Looking into the explanantia the studies referred to student characteristics, different kinds of education measures, and variations in contextual circumstances. These analyzes presented what was regarded as significant results based on the strength in associations between categories and variables – such as early differentiation in a school system is related to increased social equity or that gender gaps differ between national contexts. In the publications we often found statements by the researchers about what analyses that would be appropriate and which conclusions that are possible to make in ILSA research – e.g. in terms of associations between variables – and what *other* kinds of research that is demanded in order to make statements about cause and effect – e.g. longitudinal studies or experiments – that the ILSA studies were not able to do.

In addition, research questions and inquiries were often based on what the data bases could to offer – what information that was possible to explore or put forwards hypotheses about. To our understanding, ILSA research entails a particular kind of statistical analysis and construction of data for defining the world of education. Its comparative problematic should not be taken as self evident, according to the controversies in international comparative studies (see chapter 3), and requires scrutiny of ILSA research beyond its internal criteria of analysis. This in turn is connected to displaying relevance to statements derived from ILSA to policy-makers and professionals.

The research field of ILSA – research communication and style of reasoning

In order to capture the research field of ILSA we analyzed not only research publications but also the history of the different ILSA research programs. It is a long history, to a large extent initiated by the research community, such as the IEA, but with supranational organizations such as the OECD and EU entering the field both as supporters and as exploiters of the expanding ILSA research. We also noted academic controversies considering international comparisons in education as examples of conflicts between the humanistics and the social sciences e.g. considering what could be valid and relevant comparisons over different contexts.

Our analyses of the ILSA research field were based on the filtrations and categorisations of research publications. A large amount of the publications appeared outside the field of research communication in peer reviewed journals and another large part were commenting ILSA or presenting strategies and solutions to educational problems, rather than presenting primary research using ILSA databases. Our conclusion is that the ILSA research field in this respect is quite heterogenous, when the subjects of its research are described. This point was supported by the rather fragmented research communication structure that we captured by means of analyses of journal publications citation of articles. This fragmentation conclusion is preliminary and should be further analyzed, but so far it points to a research field having many addressees, a field in demand an improved social and intellectual organization (Whitley, 2000) in order to increase the perceived validity and coherence of research results in ILSA research.

However, there is rather a homogeneous intellectual organization of ILSA in terms of what we, based on Hacking (1992a) named style of reasoning. This refers to the ways research objects are formulated, how research inquiries are carried out, and what is considered as valid statements in this research process. Our research review – as concluded in the previous section – showed a research field with distinct characteristics. There is an internal relation in the formulation of explanandum and explanans as knowledge object plus accepted procedures for accepting or rejecting statements concerning this relation – e.g. when comparing school performances among different parts of the population. This is to our understanding basic in the style of reason at work in international large scale assessments. Such a style of reason sets limits as well as it opens for specific analyses and production of valid statements concerning the research problematic in focus. Such a style of reasoning is at the core of ILSA research and understandings its principles is needed order to capture the qualities of knowledge contributions from ILSA research. In more general terms, it is in research reviews important for research reviews to analyze what styles of reason which is at work in a field of research and by means of this define the basis for knowledge production in the current field.

Analyzing translations of ILSA research

It is vital to make a distinction between the contributions of ILSA in the research context and how these contributions are translated to other contexts with other discourses and “languages” or procedures for the making of meaningful or valid statements. This is a complex matter, where e.g. distinctions made in ILSA research are changing meaning and significance when translated into cultural and political processes.

For instance, the invention of International Large-Scale Assessments (ILSA) is related to a particular kind of cosmopolitanism (see Rabinow, 1996). This cosmopolitanism entails the use of science as a technological process to operationalize social practices that uncouples them from sedimented historical, cultural and, we should add, political processes. The OECD and TIMSS programs when placed in relation to the agencies of the *grey-zone* create systems of universal classifications that serve as comparisons among disparate social, cultural and political contexts. This “new comparative statistics” define and enforce the new social realities that remove historical and social references in the name of management efficiency as the science of progress and welfare. We argued that this approach defined context in a very particular way when referencing only to the given data sets of the ILSA. The purpose of the codifications and standardization is to represent equivalences that relate to the classifications and their utility to the given models of schools. We argued in the chapter on the *grey-zone*, for example, that these classifications are defined as the context to elaborate the management models in which dimensions of time and space are eliminated. The categories serve as elements or factors of the abstract qualities of the “*system*” from which the measurements are made.

This brings to light a number of scientific and social aspects to the ILSA which we consider to present more in detail as ten specific points:

One is that it entails a theory of schools that serve as abstraction for developing statistical categories for use in comparing. Driving the different ILSA research and the *grey-zone* are thinking of the school as systems that linguistically are thought of as embodying structures and interrelate parts that gives comprehensibility to a whole and through which the abstraction serves as a device compared across particular localities and spaces.

Second, this mode of reasoning embodies *a priori* assumptions about what the school system *should be*. The numbers that compare school system did not exist prior to the ILSA in the sense that they emerged within the creation of statistical procedures to find equivalent among different sets of data. In the ILSAs they emerge from the manner of measuring and calculating so they seem to exist as in fact what schools are, to say something about schools but also national development and modernization rather than as an abstraction. The irony of the ILSA is that what counts for inaugurating the models of intervention for the future is already imagined in the measurements that assume a consensus and stabilizing of the present.

Third, the abstraction provides the universal characteristics of school systems that can be placed on a continuum of values to differentiate national school systems. This is clear in the modeling of interventions and the descriptions of different national school systems that appear in the *grey-zone* where the results of ILSA research are used. The national school assessments are placed in a grid of other nations as a universal subject whose needs, potentialities and norms can be discovered through the assessments and its correlations with, for example, teacher, children’s and family attributes by codifications and standardization of the particular kind of science. The universalized characteristics and capacities given to the system provide the norms and distinctions to assess and organize change processes.

Fourth, the significance of the ILSA is that the seemingly practical, management question of school success is a theoretical question bound to abstractions and magnitudes that are not merely descriptive measurements of how things work. The theoretical abstraction is significant. It seeks to create a new kind of person from the correctly designed spaces and forms embodied in the norms and values of the ILSA. How else can one for instance explain the assertion of PISA that is measuring the practical knowledge of science and mathematics and the participation that students will need for the future. No scientist, mathematician or political theorist would accept this claim of knowing the future. The ILSA sciences definition of needs of schools has little to do with need but with desires that are also of the nation that the technical planner would meet.

Fifth, the ILSAs embody the hope that its knowledge would lead to concrete action in which systematic and concreted action can actualize the abstraction in the future. To achieve the articulation between the institutions and the population of schools, social facts had to be brought into a standardized grid. This process entails an

objective and objectifying vocabulary for individual and social needs as well as a functionalist understanding of education. Further, there is an intersection of macro- and micro-knowledge and powers through establishing homogeneous typologies of education that become standards to analyze needs, and setting normalizing criteria for their usage.

Sixth, a new form of expertise is formed, for instance taking the McKinsey reports or the OECD agents as examples. Norms and means are joined. Norms not only classify teachers and children but serve as the basis of intervention to hasten their creation and stabilization. Educational systems and its populations that do not qualify are not eliminated or excluded but offered the possibility of consulting with the ILSA to enter the selected normal community of the achieving school. This is the operational purpose of the OECD national reviews and McKinsey reports. It also gives the new ILSA work the right to enter into nations to check its schools and establish an obligatory system to enforce the regulation of the regular habits.

Seven, there is a particular kind of goal steering through the expertise of the ILSA. The ILSAs are not a project of regulating and ameliorating a locality (nation) and its inhabitants, but of treating both as a matter to be formed and normed so as to be acted through a thoroughly voluntarism program.

Eight, the codification and standardizing establish universalizing norms and a system of stratification that displace the prior class-based disciplinary tactics as well as a localism (the nation) in defining value.

Nine, the creation and standardization of equivalences are not merely about the teacher, and student to achievement or assessment scores. They embody cultural theses that are universalized and the capacities, capabilities, and dispositions of modes of life. We argued, for example, that the learning gaps were related to differences between expectations and performances for individual students in terms of socio-economic status, race/ethnicity and gender. These were explained, however, by pointing to universal social and psychological factors that, in effect and often through statistical measures, keep social and cultural factors as constants to identify universals such as student careers, family status, attending collect, teacher performance pay, student homework, reading performances. The classifications are administrative characteristics that can be codified and administered in a similar way across contexts. There were some differences in the TIMSS study, such as how educational expectations affected careers and attendance in higher education. But other than that, there was a continual effort in the research to identify social and psychological qualities of the child that hinders or fosters success, and on teacher knowledge of subject.

Finally, the democratic appeal given by the establishing universal categories to compare nations inscribed differences. These differences were not only organizational. The concept of "context" that was applied used the universal characteristics of children, families and communities that fostered or hindered successful school performance considered as causes of disequilibrium. The "causes" were of a practical relation rather than formal but whose effect was to define psychological and sociological distinctions that differentiated the successful from failing child, the later requiring invention and remediation. The insertion of difference entailed double gestures, or unarticulated ideas or sentiments that established universal qualities to differentiate what integrates the individual into society for its general development and progress. The consensual norms that constituted success created a scaling and hierarchy of values about the child, family and community that were recognized as different and not in harmony with system norms of effectiveness. The assumptions of disequilibrium and the pathology of the system functions, in effect, as dividing and excluding in the name of seeking inclusion and system harmony.

Considering the School Research Institute

To end, we will present four conclusions that we consider of vital importance for the School Research Institute:

- In this research review we have developed a model for a systematic research review on international comparisons of education by means of International Large Scale Assessments. This configurative model is based on the analysis of arguments by which research results are made plausible and not on research results as such. Considering the contingent and contextually informed character of education and schooling, we think this is valid model that is essential for research reviews and points to the inadequacy of focusing on results as such.

- Considering the validity and sustainability of research reviews, we argue that it is important to analyze the characteristics of the research field, where research is communicated and the validity of statements are reviewed and tested by peers. Considering the ILSA research field we draw the preliminary conclusions that it is coherent in style of reason, but fragmented in research communication and sensitive in their relevancing efforts to professional and political decision-making.
- When analyzing ILSA research it became evident that the contributions from research are based on a specific style of reasoning, that is the research objects as well as reasons for arguing that a statement or a conclusion is valid or not. In ILSA research it is a way of reasoning about a population, about categorisations and strength of associations between different variables. This style of reasoning is a prerequisite for putting forwards conclusions as well as recommendations and practical inquiries. In research reviews, like the one we did here, it is vital to clarify what style of reasoning that is structuring the knowledge contributions.
- We identified the importance to capture the translation – or rather contextualization – of educational research into education. This is a dynamic space populated by different claims of knowledge as well as organizations and experts. This was illuminated by the case of the McKinsey reports, based on PISA results but outside the peer reviewed research field. These reports were prominent publications in public discourses and having authors being highly recognized experts in policy analyses. But they had not passed scientific reviewing. Given this, an important task for the institute is to identify such grey-zone actors in-between educational research and education and teachers' professional work and to analyze how ILSA statements are translated into norms for education or distinctions of people in the population.

Based on these conclusions we identify a risk that research reviews will be legitimized by the sheer amount of publications that they are based on and not on the quality of the arguments they produce. Furthermore, it is important to capture the impact of grey-zone actors in translating research into political and professional matters. To our understanding it is vital for the School Research Institute to structure the arena of research communication where the quality of the argument is in focus and mutual respect between the participants is ensured.

A final point on the research review

To carry out a systematic research review on international comparisons of education by means of International Large Scale Assessments (ILSA) turned out to matter more of capturing and analyzing arguments and less about collating research results as such. This was partly a consequence of the current heterogeneous organization of ILSA primary research, partly due to the specific way of reasoning that we conceived of as important to understand ILSA research and its reception and translation into different educational context.

We had considerable help of search engines to identify relevant research and bibliometrics, to analyze the research field and its communicative characteristics, and from the literatures concerning systematic research reviews in the ways of organizing the review by means of a stepwise review design and using distinct criteria for screening and filtering the huge amount of research publications that are presented in ILSA.

But the main point was to find ways of dealing with arguments and knowledge contributions in this field of study. We found a workable way here by analyzing explananda and explanantia in the research and by mapping the research by means of that. When doing so we realized that we needed a tool to capture the specific ways of designing research and determining what was valid or invalid statements in ILSA research. That could have been a way of identifying “paradigms”, but we regarded as a more adequate and precise way was to use “style of reasoning”, as used by Hacking (1992a), in order to capture the research object of ILSA and the statistical reasoning that is of vital importance here. This style and its historical development is Hacking paying detailed attention to as a way of analyzing populations – e.g. means and variations over certain characteristics, and also in comparing characteristics for identified categories in populations. Such reasoning is developing over time and so does procedures for developing procedures for verification or falsification of statements. This is not to be understood as a kind of relativism. Instead, the point is that styles “... do not exist until we bring them into

being” Hacking, 1992 b, p155), which implies that we need to capture these styles at work in order to understand the knowledge contributions that is claimed within this style as well as the constraints when translating them into other contexts. This is necessary in order to assess arguments in research but it is also vital in assessing the validity of the translation of arguments into political and professional discourses!

Two discoveries in this work were of special importance for the translation point we are making here. Firstly, the share of actual primary research doing international comparisons was small. A minor part of the publications was actually presenting studies on ranking and efficiency, while a larger part ILSA research presented analyses of different kinds of educational inequalities and social and cultural biases in different educational systems. These findings were not consistent with the picture of ILSA research in public discourses, where notions of efficiency and ranking of educational systems dominated mass media.

Thus, we turned to such public discourses. Here we made the second discovery, which was the identification of a *grey-zone* where different actors translated ILSA research into seemingly relevant data for policy-making and education management. These translations were outside the research field and the style of reasoning at work there, but referred extensively on ILSA data and were actually gaining legitimacy from this referencing with the support of e.g. the OECD. Our point is not that the grey-zone is a place to avoid. Instead, there is an urgent need to analyze this grey-zone and the communication going on there between different agents having different resources and presenting knowledge claims of different kinds. In addition there is a need to understand ways in the potential making of socially robust knowledge and of pluralistic regimes of expertise (see e.g. Nowotny et al, 2003) and its societal significances.

The concern in the review to focus on the ways that arguments are structured to give relevance to school outcomes is, as we discussed, not merely an academic problem of science. The ILSA research is very important in policy initiatives and curriculum change. This makes it necessary to carefully scrutinize ILSA and what it can offer for wise and sensible policy and planning in education. Thus, it was regarded as necessary to conduct the review by analyzing the framing of research problematics that organize conceptualizations and specific practices in the research inquiries on education. An example of this is the ways that PISA research are used in putting forwards international ranking of school performances as evidence of the standard of national education qualities. Another example concerns the way “achievement gaps” are defined by a combination of taxonomic groups and testing results in ILSA research. Our point is not that such statements are invalid or un-true. Our point is that what is conceived of as valid statements about the task and functioning of education are statements framed by a specific style of reasoning at work. It is on one hand important to clarify such framing for possible knowledge claims – some conclusions are not possible. On the other hand, this style of reasoning has the potential to structure thinking about education, e.g. in terms of categorizations of students and schools as well as the tasks for curricula and educational systems. To recognize such a knowledge problematic (Popkewitz & Lindblad, 2000) with the distinctions and exclusions at work is of vital concern in the making and use of systematic research reviews.

From this point of view, research reviews are important for making apparent what is recognized through the style of reason that is organizing the data about – and solutions for – national education systems. A conceptual review, then, is to allow for public and research reflections of the arguments and not to create authorities based on the number of research publications that support a certain conclusion. To end, a research review is not only a way to summarize and organize knowledge contributions within a field of study. A research review is also a way to identify the grey-zone between research and education affairs – for better or worse. This is vital for understanding the translation of educational research in society.

LITTERATURE

- Baker, D., & LeTendre, G. K. (2005). *National differences, global similarities: World culture and the future of schooling*. Stanford University Press.
- Bunge, M. (1967). *Scientific research*, vol. I & II. New York, 1(7).
- Burkhardt, H. & Schoenfeld, A. H. (2003) Improving educational research: toward a more useful, more influential, and better funded enterprise. *Educational Researcher* 32(9) pp.3-14.
- Caplan, N. (1979). The two-communities theory and knowledge utilization. *American Behavioral Scientist* 22(3) pp.459-470.
- Carlgren, I., & Lindblad, S. (1991). On teachers' practical reasoning and professional knowledge: Considering Conceptions of Context in Teachers' Thinking. *Teaching and teacher education*, 7(5), 507-516.
- Carvalho, L. M. (2009) Production of OECD's "Programme for International Student Assessment" (PISA). Know&Pol (Knowledge and Policy in education and health sectors). European Commission within the Sixth Framework Program.
- Desrosières, A. (1991). How to make things which hold together: Social science, statistics and the state. In *Discourses on society* (pp. 195-218). Springer Netherlands.
- Domínguez, M., Vieira, M. J., & Vidal, J. (2012). The impact of the Programme for International Student Assessment on academic journals. *Assessment in Education: Principles, Policy & Practice*, 19(4), 393-409.
- Elley, W. (2002) Evaluating the impact of TIMSS-R (1999) in low- and middle-income countries: An independent report on the value of World Bank support for an international survey of achievement in mathematics and science.
- Encinas-Martin, M (2006) A Global Survey of Educational Evaluation: International, Regional, and National Assessments of Student Learning. Paper commissioned for the EFA Global Monitoring Report 2007, Strong Foundations: Early Childhood Care and Education
- Ertl, H (2006) Educational standards and the changing discourse on education: The reception and consequence of the PISA study in Germany. *Oxford Review of Education* 32(5) pp.619-634.
- Fenstermacher, G. D., & Richardson, V. (1993). The elicitation and reconstruction of practical arguments in teaching. *Journal of curriculum studies*, 25(2), 101-11
- Foshay, A. W., Thorndike, R. L., Hotyat, F., Pidgeon, D. A., & Walker, D. A. (1962) *Educational Achievements of 13 Year Olds in Twelve Countries*. UNESCO Institute of Education: Hamburg.
- Fraser, S (1964) *Jullien's plan for comparative education 1816-1817*. Teachers College, Columbia University: New York.
- Gardner, D. C. (1983) *A Nation at Risk: The Imperative for Educational Reform*. Washington: National Commission on Excellence in Education.
- Gautherin, J (1993) Marc-Antoine Jullien (Jullien de Paris) (1775-1848). *Prospects: the quarterly review of comparative education*. Vol. XXIII, No. 3/4 pp.757-73. UNESCO: Paris.
- Goldstein, H. 2004. *International comparisons of student attainment: Some issues arising*
- Gough D, Oliver S, & Thomas J (2013) *Learning from Research: Systematic Reviews for Informing Policy Decisions: A Quick Guide*. A paper for the Alliance for Useful Evidence. London: Nesta.

- Greenspan, A (2004) March 11. Address to the committee on Education and the Workforce, U.S. House of Representatives. Washington: The Federal Reserve Board.
- Grek, S. (2009). Governing by numbers: The PISA ‘effect’ in Europe. *Journal of education policy*, 24(1), 23-37.
- Grisay, A., & Monseur, C. (2007). Measuring the equivalence of item difficulty in the various versions of an international test. *Studies in Educational Evaluation*, 33(1), 69-86.
- Hacking, I. (1990). *The taming of chance* (Vol. 17). Cambridge University Press.
- Hacking, Ian (1992a) “Style” for historians and philosophers. *Studies in the History and Philosophy of Science*. 23(1): pp.1-20.
- Hacking, I. (1992b). Statistical language, statistical truth, and statistical reason: The selfauthentication of a style of scientific reasoning. *The social dimensions of science*, 3, 130-157.
- Hansen, K. Y., Gustafsson, J. E., Rosén, M., Sulkunen, S., Nissinen, K., Kupari, P., ... & Mejdning, J. (2014). *Northern Lights on TIMSS and PIRLS 2011: Differences and similarities in the Nordic countries*. Nordic Council of Ministers.
- Hanushek, E. A. & Woessmann, L. (2012) Schooling, educational achievement, and the Latin American growth puzzle. *Journal of Development Economics* 99(2) pp.497-512
- Holmes, B (1981) *Comparative Education: some consideration of method*. Allen & Unwin: London.
- Hopmann, S. T., Brinek, G., & Retzl, M. (2008). *PISA According to PISA*. Wien: LT Verlag
- Howie, S. (2011) The involvement of African countries. In: *IEA 1958-2008: 50 Years of Experiences and Memories*, Ed. C. Papanastasiou, T. Plomp and E. Papanastasiou, 317-372. Nicosia: Cultural Center of the Kykkos Monastery.
- Hultin, M. (1968) *Skolans produktivitet*. Almqvist & Wiksell: Stockholm.
- Husén, T. (1967) (Ed.) *International study of educational achievement in mathematics: A comparison of twelve countries*. Volume I and II. John Wiley & Sons: New York.
- Hutchinson, D. & Schagen, I. (2007) *Comparisons Between PISA and TIMSS – Are We the Man with Two Watches?*
(http://www.iea.nl/fileadmin/user_upload/IRC/IRC_2006/Papers/IRC2006_Hutchison_Schagen.pdf)
- Igo, S. E. (2007). *The averaged American: Surveys, citizens, and the making of a mass public*. Harvard University Press.
- International Student Assessment on academic journals. *Assessment in Education: Principles, Policy & Practice*. Vol.19 No. 4 pp.393-409.
- Jepperson, R., & Meyer, J. W. (2011). Multiple Levels of Analysis and the Limitations of Methodological Individualisms*. *Sociological Theory*, 29(1), 54-73.
- Journal of Educational Research* 54, no. 1: 37–59.
- Kamens, D H & Benavot, A (2011) National, regional and international learning assessments: trends among developing countries 1960-2009. *Globalisation, Societies and Education* Vol. 9 No. 2 pp.285-300
- Kamens, D. H. (2009) Globalization and the growth of international educational testing and national assessment. *Comparative Education Review* 54(1) pp.5-25.

- Kandel, I. L., (1933) *Comparative Education*. Houghton Mifflin: Boston.
- Keeves, J. P. (2011) IEA's contribution to research, methodology and measurement in education. In: *IEA 1958-2008: 50 Years of Experiences and Memories*, Ed. C. Papanastasiou, T. Plomp and E. Papanastasiou, 317-372. Nicosia: Cultural Center of the Kykkos Monastery.
- Keeves, J. P., (1968) *Variation in Mathematics Education in Australia*. ACER: Sydney.
- Kellaghan, T., Greaney, V., & Murray, S. (2009). *Using the results of a national assessment of educational achievement (Vol. 5)*. World Bank Publications.
- Kuusinen, J. (1967) *School Systems and Level of Achievement: Some Notions on the Basis of the IEA Study in Mathematics*. University of Finland: Helsinki.
- Landahl, J & Lundahl, C (2013) (Mis-) Trust in Numbers: shape shifting and directions in the modern history of data in Swedish educational reform. In Lawn, M (2013) (Ed.) *The Rise of Data in Education Systems: collection, visualization and use*. Comparative Histories of Education, Symposium Books: Oxford.
- Larsen Sogaard, M, Kornbeck Pihl, K, Kristensen Müller, R, Larsen Rode, M & Sommersel Bjornoy, H (2013) *Dropout Phenomena at Universities: What is Dropout? Why does Dropout Occur? What Can be Done by the Universities to Prevent or Reduce it? A systematic review*. Danish Clearinghouse for Educational Research, Aarhus University: Copenhagen.
- Lawn, M (2013) *The Internationalization of Education Data: exhibitions, tests, standards and associations*. In Lawn, M (2013) (Ed.) *The Rise of Data in Education Systems: collection, visualization and use*. Comparative Histories of Education, Symposium Books: Oxford.
- Lehmann, R. (2011) The impact of IEA on educational policy making in Germany. In: *IEA 1958-2008: 50 Years of Experiences and Memories*, Ed. C. Papanastasiou, T. Plomp and E. Papanastasiou, 317-372. Nicosia: Cultural Center of the Kykkos Monastery.
- Lockheed, M. (2009) *The Effects of the Program of the Assessment of Student Achievement (PASA) on the Educational Assessment Capacity in Developing Countries: An Independent Evaluation*. Washington: The World Bank.
- Lussi Borer, V & Lawn, M (2013) *Governing Education Systems by Shaping Data: from the past to the present, from national to international perspectives*. *European Educational Research Journal* Vol. 12 No. 1 pp. 48-52
- Mann, H., (1844) *Seventh Annual Report of the Secretary of the Board*. Dutton & Wentworth: Boston.
- Martens, K. (2007), "How to become an influential actor – The "comparative turn" in OECD education policy", in K. Martens, A. Rusconi & K. Lutz, (eds.), *Transformations of the state and global governance*. London: Routledge, 40-56.
- Mullis, I. V. & Martin, M. O. (2007) *Lessons learned: What international assessments tell us about math achievement*. In: *TIMSS in Perspective: Lessons Learned from IEA's Four Decades of International Mathematics Assessments*. Washington: Brooking Institution Press.
- Mullis, I. V., Martin, M. O., Gonzales, E. J. (2004) *TIMSS International Mathematics Report. Findings from IEA's Trends in International Mathematics Assessments*. Washington: Brooking Institution Press.
- Noah, H. & Eckstein, M., (1969) *Toward a Science of Comparative Education*. Macmillan: London.
- Nóvoa, A. & Lord, M. (Ed.) (2002), *Fabricating Europe: The formation of an education space*. Dordrecht/Boston/London: Kluwer Academic Publishers.

- Nóvoa, A. & Yariv-Mashal, T., (2003) Comparative Research in Education: a mode of governance or a historical journey? *Comparative Education* Vol. 39, No. 4 pp.423-438.
- Nowotny, H., Scott, P. & Gibbons, M. (2003). Introduction: Mode 2'revisited: The new production of knowledge. *Minerva*, 41(3), 179-194.
- Oakley A et al (2005) The politics of evidence and methodology: lessons from the EPPI-Centre. *Evidence & Policy* 1 (1): 5–31.
- OECD (1999) *Measuring Student Knowledge and Skills: A New Framework for Assessment*. Paris: OECD.
- OECD (2001) *Knowledge and skills for life: First results of Programme for International Student Assessment*. Paris: OECD.
- OECD (2006) OECD (Organization for Economic Cooperation and Development). 2006. *Education at a Glance, 2006*. Paris: OECD.
- Owens, T. L. (2013) Thinking Beyond League Tables: a review of key PISA research questions. *PISA, Power and Policy : the emergence of global educational governance*. Ed. H-D. Meyer and A. Benavot. Oxford: Oxford Studies in Comparative Education; 23:1. Symposium Books.
- Pettersson, D (2008) Internationell kunskapsbedömning som inslag i nationell styrning av skolan. *Acta Universitatis Upsaliensis. Uppsala Studies in Education* No 120. Uppsala universitet: Uppsala.
- Pettersson, D (2014) Three Narratives: National Interpretations of PISA. *Knowledge Cultures*.2(4) (edited by David Cole & Christine Woodrow).
- Petticrew, M., & Roberts, H. (2003). Evidence, hierarchies, and typologies: horses for courses. *Journal of epidemiology and community health*, 57(7), 527-529.
- Pidgeon, D. A., (1967) *Achievement in Mathematics: A National Study in Secondary Schools*. NFER: Upton Park.
- Poovey, M. (1998). *A history of the modern fact: Problems of knowledge in the sciences of wealth and society*. University of Chicago Press.
- Popkewitz, T. (2014). Social Epistemology, the Reason of “Reason” and the Curriculum Studies. Special Issue: Nuevas Perspectivas sobre el Currículum Escolar. *Education Policy Analysis Archives* DOI: <http://dx.doi.org/10.14507/epaa.v22n22.2014>
- Porter, T. M., (1995) *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton University Press: Princeton, New Jersey.
- Postlethwaite, N & Kellaghan, T (2008) National assessments of educational achievement. The International Institute for Educational Planning (IIEP): Paris
- Postlethwaite, N & Ross, K N (1992) *Effective schools in reading: implications for educational planners. An exploratory study*. IEA: The Hague
- Prøitz, T. (2015) Metoder for systematiske kunnskapsoversikter – relevant, tilgjengelig og praktisk anvendbar? Vetenskapsrådets publikasjonsserie, forthcoming spring 2015
- Purves, A. C., (1987) The Evolution of the IEA: A Memoir. *Comparative Education Review*, Vol. 31, No. 1 pp.10-28.
- Rabinow, (1991): *On the archaeology of late modernity, essays on the anthropology of reason*. In Rabinow, P. The foucault reader (p. 244). London: Penguin.

- Rizvi, F. & Lingard, B. (2006), Globalisation and the changing nature of the OECD's educational work, in Lauder, H., Brown, P., Dillabough, J. & Halsey, A. H. (eds.), *Education, globalization and social change*. Oxford: Oxford University Press, 129-145.
- Robitaille, D. B. (2000) *The Impact of TIMSS on the Teaching and Learning of Mathematics and Science*. Vancouver: Pacific Educational Press.
- Rose, N. (1999). *Powers of freedom: Reframing political thought*. Cambridge university press.
- Sadeh, T (2013) *From Search to Discovery*. IFLA WLIC: Singapore.
- Saito, M. C. & Capele, F. (2010) *Monitoring the quality of education: Exploration of concept, methodology, and the link between research and policy*. In: *The Impact of International Achievement Studies on National Education Policymaking: International Perspectives on Education and Society*. Bingley: Emerald Publishing.
- Schleicher, A & Saito, M (1995) *Data building and data management*. IIEP: Paris
- Schwippert, K & Lenkeit, J (2012) *Progress in reading literacy in national and international context: The impact of PIRLS 2006 in 12 countries*. Münster: Waxman.
- Schwippert, K. (2007) *Progress in Reading Literacy: The Impact of PIRLS 2001 in 13 Countries*. Hamburg: Waxman.
- Smyth, (2008) *The Origins of the International Standard Classification of Education*. *Peabody Journal of Education*. Vol. 83, No. 1 pp.5-40.
- Tatto , M. T., Schwillie, J., Senk, S. L., Ingvarson, L., Rowley, G., Peck, R., Bankov, K., Rodirguez, M, & Rekase, M. (2012) *Policy, Practice and Readiness to Teach Primary and Secondary Mathematics in 17 Countries: Findings from the IEA Teacher Education and Development Study in Mathematics (TEDS-M)*. Amsterdam: IEA.
- Tuijnman, A. C. & Postlethwaite, N, (1994) *Monitoring the Standards of Education*. Oxford: Pergamon.
- von Wright, G. H. (1983): *Practical Reason*. Blackwell (1983)
- von Wright, G. H. (2004). *Explanation and understanding*. Cornell University Press.
- Wagemaker, H. (2011) *IEA: International studies, impact and transition*. In: *IEA 1958-2008: 50 Years of Experiences and Memories*, Ed. C. Papanastasiou, T. Plomp and E. Papanastasiou, 317-372. Nicosia: Cultural Center of the Kykkos Monastery.
- Wagemaker, H. (2014)*International Large-Scale Assessments: From Research to Policy*. In: *Handbook of International Large-Scale Assessment: Background, Technical Issues, and Methods of Data Analysis*. Eds: L. Rutkowski, M. van Davier & D. Rutkowski. Boca Raton, London & New York: CRC Press, Taylor & Francis Group.
- Walker, D.A. (1962)'An analysis of the reactions of Scottish teachers and pupils to items in the geography, mathematics and science items', in *Educational Achievements of Thirteen-year-olds in Twelve Countries*, Hamburg: UNESCO Institute for Education.
- World Bank (2005) *Expanding opportunities and building competencies for young people*. World Bank: Washington DC
- World Bank, (2011a) *Learning for all: Investing in People's Knowledge and Skills to Promote Development: Education Sector Strategy 2020*. Washington: World Bank.

World Bank, (2011b) The Arab Regional Agenda on Improving Education Quality (ARIEQ). Tunis: World Bank.

Wu, M (2009) A Critical Comparison of the Contents of PISA and TIMSS Mathematics Assessments (https://edsurveys.rti.org/PISA/documents/WuA_Critical_Comparison_of_the_Contents_of_PISA_and_TIMSS_psg_WU_06.1.pdf)

MAPPED ARTICLES

Overview of the mapped articles (including PISA, TIMSS and CIVED/ICCS)

- Ainley, M. & Ainley, J. (2011) A Cultural Perspective on the Structure of Student Interest in Science. *International Journal of Science Education*, Vol. 33 Issue 1, p51-71.
- Alegre, M. A. & Ferrer, G. (2010) School regimes and education equity: some insights based on PISA 2006. *British Educational Research Journal*. Vol. 36 Issue 3, p433-461.
- Alviar-Martin, T., Randall, J. D., Usher, E. L. & Engelhard, G. (2008) Teaching Civic Topics in Four Societies: Examining National Context and Teacher Confidence. *Journal of Educational Research*, Vol. 101 Issue 3, p177-188.
- Akiba, M., LeTendre, G. K., Baker, D. P. & Goesling, B. (2002) Student Victimization: National and School System Effects on School Violence in 37 Nations. *American Educational Research Journal*. Vol. 39 Issue 4, p829-853.
- Akiba, M., LeTendre, G. K. & Scribner, J. P. (2007) Teacher Quality, Opportunity Gap, and National Achievement in 46 Countries. *Educational Researcher*, Vol. 36 Issue 7, p369-387.
- Ammermüller, A., Heijke, H. & Wößmann, L. (2005) Schooling quality in Eastern Europe: Educational production during transition. *Economics of Education Review*. Vol. 24 Issue 5, p579-599.
- Baker, D. P., Akiba, M., LeTendre, G. K. & Wiseman, A. W. (2001) Worldwide Shadow Education: Outside-School Learning, Institutional Quality of Schooling, and Cross-National Mathematics Achievement. *Educational Evaluation and Policy Analysis*. Vol. 23 Issue 1, p1-17.
- Barber, C., Fennelly, K. & Torney-Purta, J. (2013) Nationalism and Support for Immigrants' Rights Among Adolescents in 25 Countries. *Applied Developmental Science*, Vol. 17 Issue 2, p60-75.
- Barber, N. (2006) Is the Effect of National Wealth on Academic Achievement Mediated by Mass Media and Computers? *Cross-Cultural Research*. Vol. 40 Issue 2, p130-151.
- Basl, J. (2011) Effect of School on Interest in Natural Sciences: A comparison of the Czech Republic, Germany, Finland, and Norway based on PISA 2006. *International Journal of Science Education*, Vol. 33 Issue 1, p145-157.
- Bielinski, J. & Davison, M. L. (2001) A Sex Difference by Item Difficulty Interaction in Multiple-Choice Mathematics Items Administered to National Probability Samples. *Journal of Educational Measurement*. Vol. 38 Issue 1, p51-77.
- Birenbaum, M., Tatsuoka, C. & Tao X. (2005) Large-scale diagnostic assessment: comparison of eighth graders' mathematics performance in the United States, Singapore and Israel. *Assessment in Education: Principles, Policy & Practice*, Vol. 12 Issue 2, p167-181.
- Bos, K. & Kuiper, W. (1999) Modelling TIMSS Data in a European Comparative Perspective: Exploring Influencing Factors on Achievement in Mathematics in Grade 8. *Educational Research & Evaluation*. Vol. 5 Issue 2, p157-180.
- Bratti, M., Checchi, D. & Filippin, A. (2011) Should you compete or cooperate with your schoolmates? *Education Economics*. Vol. 19 Issue 3, p275-289.
- Brozo, W. G., Shiel, G. & Topping, K. (2008) Engagement in reading: Lessons learned from three PISA countries. *Journal of Adolescent & Adult Literacy*, Vol. 51 Issue 4, p304-315.

- Buccheri, G., Gurber, N. A. & Bruhwiler, C. (2011) The Impact of Gender on Interest in Science Topics and the Choice of Scientific and Technical Vocations. *International Journal of Science Education*, Vol. 33 Issue 1, p159-178.
- Bybee, R. & McCrae, B. (2011) Scientific Literacy and Student Attitudes: Perspectives from PISA 2006 science. *International Journal of Science Education*, Vol. 33 Issue 1, p7-26.
- Ce, S. (2005) How American Middle Schools Differ From Schools of Five Asian Countries: Based on cross-national data from TIMSS 1999. *Educational Research & Evaluation*. Vol. 11 Issue 2, p179-199.
- Cheng, S. K. & Seng, Q. K. (2001) Sex Differences in TIMSS Mathematics Achievement of Four Asian Nations. *Studies in Educational Evaluation*, Vol. 27 Issue 4, p331-340.
- Cho, I. (2012) The effect of teacher–student gender matching: Evidence from OECD countries. *Economics of Education Review*. Vol. 31 Issue 3, p54-67.
- Chudgar, A. & Luschei, T. F. (2009) National Income, Income Inequality, and the Importance of Schools: A Hierarchical Cross-National Comparison. *American Educational Research Journal*, Vol. 46 Issue 3, p626-658.
- Cogan, L. S. & Schmidt, W. H. (2002) “Culture Shock” – Eighth-Grade Mathematics From an International Perspective. *Educational Research & Evaluation*. Vol. 8 Issue 1, p13-39.
- Desimone, L. M., Smith, T., Baker, D. & Ueno, K. (2005) Assessing Barriers to the Reform of U.S. Mathematics Instruction from an International Perspective. *American Educational Research Journal*, Vol. 42 Issue 3, p. 501-535.
- Detmers, S., Trautwein, U. & Ludtke, O. (2009) The relationship between homework time and achievement is not universal: evidence from multilevel analyses in 40 countries. *School Effectiveness & School Improvement*, Vol. 20 Issue 4, p375-405.
- Dodeen, H., Abdelfattah, F., Shumrani, S. & Hilal, M. A. (2012) The Effects of Teachers’ Qualifications, Practices, and Perceptions on Student Achievement in TIMSS Mathematics: A Comparison of Two Countries. *International Journal of Testing*. Vol. 12 Issue 1, p61-77.
- Dronkers, J. & Robert, P. (2008) Differences in Scholastic Achievement of Public, Private Government-Dependent, and Private Independent Schools: A Cross-National Analysis. *Educational Policy*, Vol. 22 Issue 4, p541-577.
- Dronkers, J., van der Velden, R. & Dunne, A. (2012) Why Are Migrant Students Better off in Certain Types of Educational Systems or Schools than in Others? *European Educational Research Journal*, Vol. 11 Issue 1, p11-44.
- Dronkers, J. & Kornder, N. (2014) Do migrant girls perform better than migrant boys? Deviant gender differences between the reading scores of 15-year-old children of migrants compared to native pupils. *Educational Research & Evaluation*. Vol. 20 Issue 1, p44-66.
- Dupriez, V., Dumay, X. & Vause, A. (2008) How Do School Systems Manage Pupils' Heterogeneity? *Comparative Education Review*, Vol. 52 Issue 2, p245-273.
- Duru-Bellat, M. & Suchaut, B. (2005) Organisation and Context, Efficiency and Equity of Educational Systems: what PISA tells us. *European Educational Research Journal*, Vol. 4 Issue 3, p181-194.
- Elchardus, M., Franck, E., DeGroof, S. & Kavadias, D. (2013) The Acceptance of the Multicultural Society Among Young People. A Comparative Analysis of the Effect of Market-Driven Versus Publicly Regulated Educational Systems. *European Sociological Review*, Vol. 29 Issue 4, p767-779.

- Else-Quest, N. M., Hyde, J. S. & Linn, M. C. (2010) Cross-National Patterns of Gender Differences in Mathematics: A Meta-Analysis. *Psychological Bulletin*. Vol. 136 Issue 1, p103-127.
- Fredriksson, U., Holzer, T., McCluskey-Cavin, H. & Taube, K. (2009) Strengths and Weaknesses in the Swedish and Swiss Education Systems: a comparative analysis based on PISA data. *European Educational Research Journal*, Vol. 8 Issue 1, p54-68.
- Gamboa, L. F. & Waltenberg, F. D. (2012) Inequality of opportunity for educational achievement in Latin America: Evidence from PISA 2006–2009. *Economics of Education Review*. Vol. 31 Issue 5, p694-708.
- Givvin, K. B., Hiebert, J., Jacobs, J. K., Hollingsworth, H. & Gallimore, R. (2005) Are There National Patterns of Teaching? Evidence from the TIMSS 1999 Video Study. *Comparative Education Review*. Vol. 49 Issue 3, p311-343.
- Hampden-Thompson, G. (2009) Are two better than one? A comparative study of achievement gaps and family structure. *Compare: A Journal of Comparative & International Education*. Vol. 39 Issue 4, p513-529.
- Hanushek, E. A. & Luque, J. A. (2003) Efficiency and equity in schools around the world. *Economics of Education Review*. Vol. 22 Issue 5, p481-502.
- Hauger, J. B. & Sireci, S. G. (2008) Detecting Differential Item Functioning across Examinees Tested in Their Dominant Language and Examinees Tested in a Second Language. *International Journal of Testing*, Vol. 8 Issue 3, p237-250.
- Helbig, M. (2012) Boys do not benefit from male teachers in their reading and mathematics skills: empirical evidence from 21 European Union and OECD countries. *British Journal of Sociology of Education*, Vol. 33 Issue 5, p661-677.
- Heuveline, P., Hongxing, Y. & Timberlake, J. M. (2010) It Takes a Village (Perhaps a Nation): Families, States, and Educational Achievement. *Journal of Marriage & Family*. Vol. 72 Issue 5, p1362-1376.
- Hooghe, M. & Dassonneville, R. (2013) Voters and Candidates of the Future: The Intention of Electoral Participation among Adolescents in 22 European Countries. *Young*, Vol. 21 Issue 1, p1-28.
- Hoskins, B., Barber, L., Van Nijlen, D., & Villalba, E. (2011) Comparing Civic Competence among European Youth: Composite and Domain-Specific Indicators Using IEA Civic Education Study Data. *Comparative Education Review*. Vol. 55 Issue 1, p82-110.
- Hoskins, B., Janmaat, J. G. & Villalba, E. (2012) Learning citizenship through social participation outside and inside school: an international, multilevel study of young people's learning of citizenship. *British Educational Research Journal*. Vol. 38 Issue 3, p419-446.
- House J. D. (2006) Mathematics beliefs and achievement of elementary school students in Japan and the United States: results from the Third International Mathematics and Science Study. *Journal of Genetic Psychology* Vol. 167 Issue 1, p31-45.
- House, J. D. & Telese, J. A. (2008) Relationships between student and instructional factors and algebra achievement of students in the United States and Japan: an analysis of TIMSS 2003 data. *Educational Research & Evaluation*. Vol. 14 Issue 1, p101-112.
- Hyunjoon P. (2008) The Varied Educational Effects of Parent-Child Communication: A Comparative Study of Fourteen Countries. *Comparative Education Review*. Vol. 52 Issue 2, p219-244.
- Im, S. & Park, H. J. (2010) A comparison of US and Korean students' mathematics skills using a cognitive diagnostic testing method: linkage to instruction. *Educational Research & Evaluation*. Vol. 16 Issue 3, p287-301.

- Isac, M. M., Maslowski, R. & van der Werf, G. (2011) Effective civic education: an educational effectiveness model for explaining students' civic knowledge. *School Effectiveness & School Improvement*, Vol. 22 Issue 3, p313-333.
- Isac, M. M., Maslowski, R., Creemers, B. & van der Werf, G. (2014) The contribution of schooling to secondary-school students' citizenship outcomes across countries. *School Effectiveness & School Improvement*, Vol. 25 Issue 1, p29-63.
- Janmaat, J. G. (2011) Ability grouping, segregation and civic competences among adolescents. *International Sociology*, Vol. 26 Issue 4, p455-482.
- Janmaat, J. G. (2012) The Effect of Classroom Diversity on Tolerance and Participation in England, Sweden and Germany. *Journal of Ethnic & Migration Studies*, Vol. 38 Issue 1, p21-39.
- Jenkins, S. P., Micklewright, J. & Schnepf, S. V. (2008) Social segregation in secondary schools: how does England compare with other countries? *Oxford Review of Education*. Vol. 34 Issue 1, p21-37.
- Jerrim, J. (2014) The Unrealistic Educational Expectations of High School Pupils: Is America Exceptional? *Sociological Quarterly*. Vol. 55 Issue 1, p196-231.
- Jerrim, J. & Choi, Á. (2014) The mathematics skills of school children: how does England compare to the high-performing East Asian jurisdictions? *Journal of Education Policy*. Vol. 29 Issue 3, p349-376.
- Kelly, D. L. (2002) The TIMSS 1995 International Benchmarks of Mathematics and Science Achievement: Profiles of World Class Performance at Fourth and Eighth Grades. *Educational Research & Evaluation*. Vol. 8 Issue 1, p41-54.
- Kennedy, K. J. (2007) Student Constructions of 'Active Citizenship': What Does Participation Mean to Students? *British Journal of Educational Studies*, Vol. 55 Issue 3, p304-324.
- Kennedy, K. J., Kuang, X. & Chow, J. K. F. (2013) Exploring Asian Students' Citizenship Values and Their Relationship to Civic Knowledge and School Participation. *Educational Psychology*, Vol. 33 Issue 3 p233-254.
- Kjaernsli, M. & Lie, S. (2004) PISA and scientific literacy: similarities and differences between the Nordic countries. *Scandinavian Journal of Educational Research*. Vol. 48 Issue 3, p271-286.
- Kjaernsli, M. & Lie, S. (2008) Country profiles of scientific competence in TIMSS 2003. *Educational Research & Evaluation*. Vol. 14 Issue 1, p73-85.
- Kjærnsli, M. & Lie, S. (2011) Students' Preference for Science Careers: International comparisons based on PISA 2006. *International Journal of Science Education*, Vol. 33 Issue 1, p121-144.
- Lauglo, J. (2013) Do more knowledgeable adolescents have more rationally based civic attitudes? Analysis of 38 countries. *Educational Psychology*. Vol. 33 Issue 3, p255-275.
- LaFontaine, D. & Monseur, C. (2009) Gender Gap in Comparative Studies of Reading Comprehension: to what extent do the test characteristics make a difference? *European Educational Research Journal*, Vol. 8 Issue 1, p69-79.
- Le Donné, N. (2014) European Variations in Socioeconomic Inequalities in Students' Cognitive Achievement: The Role of Educational Policies. *European Sociological Review*. Vol. 30 Issue 3, p329-343.
- Lee, B. (2014) The influence of school tracking systems on educational expectations: a comparative study of Austria and Italy. *Comparative Education*. Vol. 50 Issue 2, p206-228.

- Leung, F. K. S. (2002) Behind the High Achievement of East Asian Students. *Educational Research & Evaluation*. Vol. 8 Issue 1, p87-108.
- Leung, F. K. S. (2005) Some Characteristics of East Asian Mathematics Classrooms Based on Data from the Timss 1999 Video Study. *Educational Studies in Mathematics*, Vol. 60 Issue 2, p199-215.
- Li, H., Lei, P-W. & Pace, C. L. (2013) Reading subskill differences between students in Shanghai-China and the US: evidence from PISA 2009. *Educational Research & Evaluation*. Vol. 19 Issue 6, p490-509.
- Liem, G. A. D., Martin, A. J., Anderson, M., Gibson, R. & Sudmalis, D. (2014) The Role of Arts-Related Information and Communication Technology Use in Problem Solving and Achievement: Findings From the Programme for International Student Assessment. *Journal of Educational Psychology*. Vol. 106 Issue 2, p348-363.
- Lin, A. R. (2014) Examining Students' Perception of Classroom Openness as a Predictor of Civic Knowledge: A Cross-National Analysis of 38 Countries. *Applied Developmental Science*, Vol. 18 Issue 1, p17-30.
- Linnakyla, P., Malin, A. & Taube, K. (2004) Factors behind Low Reading Literacy Achievement. *Scandinavian Journal of Educational Research*, Vol. 48 Issue 3, p231-249.
- Liu, O. L. & Wilson, M. (2009) Gender Differences and Similarities in PISA 2003 Mathematics: A Comparison between the United States and Hong Kong. *International Journal of Testing*. Vol. 9 Issue 1, p20-40.
- Ma, X. (2008) Within-School Gender Gaps in Reading, Mathematics, and Science Literacy. *Comparative Education Review*, Vol. 52 Issue 3, p437-460.
- Maiello, C., Oser, F. & Biedermann, H. (2003) Civic Knowledge, Civic Skills and Civic Engagement. *European Educational Research Journal*, Vol. 2 Issue 3, p384-395.
- Marks, G. (2006) Are between- and within-school differences in student performance largely due to socio-economic background? Evidence from 30 countries. *Educational Research*. Vol. 48 Issue 1, p21-40.
- Marks, G. (2008) Accounting for the gender gaps in student performance in reading and mathematics: evidence from 31 countries. *Oxford Review of Education*. Vol. 34 Issue 1, p89-109.
- Martins, L. & Veiga, P. (2010) Do inequalities in parents' education play an important role in PISA students' mathematics achievement test score disparities? *Economics of Education Review*. Vol. 29 Issue 6, p1016-1033.
- McDaniel, A. (2010) Cross-National Gender Gaps in Educational Expectations: The Influence of National-Level Gender Ideology and Educational Systems. *Comparative Education Review*. Vol. 54 Issue 1, p27-50.
- Mei-Shiu C. (2008) Achievements and self-concepts in a comparison of math and science: exploring the internal/external frame of reference model across 28 countries. *Educational Research & Evaluation*. Vol. 14 Issue 3, p235-254.
- Mei-Shiu C. (2010) Effects of science interest and environmental responsibility on science aspiration and achievement: gender differences and cultural supports. *Educational Research & Evaluation*. Vol. 16 Issue 4, p345-370.
- Menezes, I. (2003) Participation Experiences and Civic Concepts, Attitudes and Engagement: implications for citizenship education projects. *European Educational Research Journal*, Vol. 2 Issue 3, p430-445.
- Merry, J. J. (2013) Tracing the U.S. Deficit in PISA Reading Skills to Early Childhood: Evidence from the United States and Canada. *Sociology of Education*. Vol. 86 Issue 3, p234-252.

- Mesic, V. (2012) Identifying Country-Specific Cultures of Physics Education: A differential item functioning approach. *International Journal of Science Education*, Vol. 34 Issue 16, p2483-2500.
- Mintrop, H. (2003) The Old and New Face of Civic Education: expert, teacher, and student views. *European Educational Research Journal*, Vol. 2 Issue 3, p446-454.
- Montt, G. (2011) Cross-national Differences in Educational Achievement Inequality. *Sociology of Education*. Vol. 84 Issue 1, p49-68.
- Mostafa, T. (2010) Decomposing inequalities in performance scores: the role of student background, peer effects and school characteristics. *International Review of Education / Internationale Zeitschrift für Erziehungswissenschaft*, Vol. 56 Issue 5/6, p567-589.
- Neuschmidt, O., Barth, J. & Hastedt, D. (2008) Trends in gender differences in mathematics and science (TIMSS 1995–2003). *Studies in Educational Evaluation*. Vol. 34 Issue 2, p56-72.
- Nixon, R. S. & Barth, K. N. (2014) A Comparison of TIMSS Items Using Cognitive Domains. *School Science & Mathematics*, Vol. 114 Issue 2, p65-75.
- Nonoyama-Tarumi, Y. & Willms, J. D. (2010) The relative and absolute risks of disadvantaged family background and low levels of school resources on student literacy. *Economics of Education Review*, Vol. 29 Issue 2, p214-224.
- O'Dwyer, L. M. (2005) Examining the Variability of Mathematics Performance and its Correlates Using Data From TIMSS '95 and TIMSS '99. *Educational Research & Evaluation*. Vol. 11 Issue 2, p155-177.
- Olande, O. (2014) Graphical artefacts: Taxonomy of students' response to test items. *Educational Studies in Mathematics*, Vol. 85 Issue 1, p53-74.
- Olsen, R. V. & Lie, S. (2011) Profiles of Students' Interest in Science Issues around the World: Analysis of data from PISA 2006. *International Journal of Science Education*, Vol. 33 Issue 1, p97-120.
- Osler, A. & Starkey, H. (2001) Citizenship Education and National Identities in France and England: inclusive or exclusive? *Oxford Review of Education*. Vol. 27 Issue 2, p287-305.
- Papanastasiou, C. & Papanastasiou, E. (2004) Major Influences on Attitudes Toward Science. *Educational Research & Evaluation*. Vol. 10 Issue 3, p239-257.
- Papanastasiou, E. (2002) Factors that Differentiate Mathematics Students in Cyprus, Hong Kong, and the USA. *Educational Research & Evaluation*. Vol. 8 Issue 1, p129-146.
- Papanastasiou, E. (2003) Science Literacy by Technology by Country: USA, Finland and Mexico. Making sense of it all. *Research in Science & Technological Education*, Vol. 21 Issue 2, p229-241.
- Papanastasiou, E., Zembylas, M. & Vrasidas, C. (2005) An Examination of the PISA Database to Explore the Relationship Between Computer Use and Science Achievement. *Educational Research & Evaluation*. Vol. 11 Issue 6, p529-543.
- Penner, A. M. (2008) Gender Differences in Extreme Mathematical Achievement: An International Perspective on Biological and Social Factors. *American Journal of Sociology*, Vol. 114, p138-170.
- Perry, L. (2009) Characteristics of Equitable Systems of Education: A Cross-National Analysis. *European Education*. Vol. 41 Issue 1, p79-100.
- Philippou, G. N. & Christou, C. (1999) Teachers' Conceptions of Mathematics and Students' Achievement: A Cross-Cultural Study Based on Results from TIMSS. *Studies in Educational Evaluation*, Vol. 25 Issue 4, p379-398.

- Pong, S-L. & Pallas, A. (2001) Class Size and Eighth-Grade Math Achievement in the United States and Abroad. *Educational Evaluation and Policy Analysis*. Vol. 23 Issue 3, p251-273.
- Post, D. & Pong, S-L. (2000) Employment during Middle School: The Effects on Academic Achievement in the U. S. and Abroad. *Educational Evaluation and Policy Analysis*. Vol. 22 Issue 3, p273-298.
- Quintelier, E. & Hooghe, M. (2013) The relationship between political participation intentions of adolescents and a participatory democratic climate at school in 35 countries. *Oxford Review of Education*. Vol. 39 Issue 5, p567-589.
- Reilly, D. (2012) Gender, Culture, and Sex-Typed Cognitive Abilities. *PLoS ONE*, Vol. 7 Issue 7, p1-16.
- Robert, P. (2010) Social origin, school choice, and student performance. *Educational Research & Evaluation*. Vol. 16 Issue 2, p107-129.
- Rutkowski, L., Rutkowski, D. & Plucker, J. (2012) International determinants of private school attendance. *Educational Research & Evaluation*. Vol. 18 Issue 6, p375-391.
- Shapira, M. (2012) An Exploration of Differences in Mathematics Attainment among Immigrant Pupils in 18 OECD Countries. *European Educational Research Journal*, Vol. 11 Issue 1, p68-95.
- Shen, C. (2002) Revisiting the Relationship Between Students' Achievement and their Self-perceptions: a cross-national analysis based on TIMSS 1999 data. *Assessment in Education: Principles, Policy & Practice*, Vol. 9 Issue 2, p161-184.
- Shen, C. & Pedulla, J. J. (2000) The Relationship between Students' Achievement and their Self-perception of Competence and Rigour of Mathematics and Science: a cross-national analysis. *Assessment in Education: Principles, Policy & Practice*, Vol. 7 Issue 2, p237-253.
- Shen, C. & Tam, H. P. (2008) The paradoxical relationship between student achievement and self-perception: a cross-national analysis based on three waves of TIMSS data. *Educational Research & Evaluation*. Vol. 14 Issue 1, p87-100.
- Schiller, K. S., Khmelkov, V. T. & Wang, X-Q. (2002) Economic Development and the Effects of Family Characteristics on Mathematics Achievement. *Journal of Marriage & Family*. Vol. 64 Issue 3, p730-742.
- Sikora, J. & Pokropek, A. (2012) Intergenerational Transfers of Preferences for Science Careers in Comparative Perspective. *International Journal of Science Education*, Vol. 34 Issue 16, p2501-2527.
- Sikora, J. & Pokropek, A. (2012) Gender Segregation of Adolescent Science Career Plans in 50 Countries. *Science Education*, Vol. 96 Issue 2, p234-264.
- Song, S., Perry, L. B. & McConney, A. (2014) Explaining the achievement gap between Indigenous and non-Indigenous students: an analysis of PISA 2009 results for Australia and New Zealand. *Educational Research & Evaluation*. Vol. 20 Issue 3, p178-198.
- Stoet, G. & Geary, D. C. (2013) Sex Differences in Mathematics and Reading Achievement Are Inversely Related: Within- and Across-Nation Assessment of 10 Years of PISA Data. *PLoS ONE*, Vol. 8 Issue 3, p1-10.
- Tarr, J. E., Uekawa, K., Mittag, K. C. & Lennex, L. (2000) A Comparison of Calculator Use in Eighth-Grade Mathematics Classrooms in the United States, Japan, and Portugal: Results From the Third International Mathematics and Science Study. *School Science & Mathematics*, Vol. 100 Issue 3, p139-151.
- Tatsuoka, K. K., Corter, J. E. & Tatsuoka, C. (2004) Patterns of Diagnosed Mathematical Content and Process Skills in TIMSS-R across a Sample of 20 Countries. *American Educational Research Journal*, Vol. 41 Issue 4, p901-926.

- Torney-Purta, J. (2002) The School's Role in Developing Civic Engagement: A Study of Adolescents in Twenty-Eight Countries. *Applied Developmental Science*, Vol. 6 Issue 4, p203-212.
- Torney-Purta, J., Wilkenfeld, B. & Barber, C. (2008) How Adolescents in 27 Countries Understand, Support, and Practice Human Rights. *Journal of Social Issues*. Vol. 64 Issue 4, p857-880.
- Torney-Purta J. & Barber C. (2011) Fostering young people's support for participatory human rights through their developmental niches. *The American Journal of Orthopsychiatry* Vol. 81 Issue 4, p473-481.
- Tramonte, L. & Willms, J. D. (2010) Cultural capital and its effects on education outcomes. *Economics of Education Review*. Vol. 29 Issue 2, p200-213.
- Turmo, A. (2004) Scientific literacy and socio-economic background among 15-year-olds--a Nordic perspective. *Scandinavian Journal of Educational Research*. Vol. 48 Issue 3, p287-305.
- Verhaegen, S., Hooghe, M. & Meeusen, C. (2013) Opportunities to learn about Europe at school. A comparative analysis among European adolescents in 21 European member states. *Journal of Curriculum Studies*. Vol. 45 Issue 6, p838-864.
- Vlaardingerbroek, B. & Neil Taylor, T.G. (2003) Teacher education variables as correlates of primary science ratings in thirteen TIMSS systems. *International Journal of Educational Development*. Vol. 23 Issue 4, p429-439.
- Waldfoegel, J. & Zhai, F. (2008) Effects of public preschool expenditures on the test scores of fourth graders: evidence from TIMSS. *Educational Research & Evaluation*. Vol. 14 Issue 1, p9-28.
- Wang, J. & Zhu, C. (2003) An In-depth Analysis of Achievement Gaps Between Seventh and Eighth Grades in the TIMSS Database. *School Science & Mathematics*, Vol. 103 Issue 4, p186-193.
- Williams, T., Williams, K., Kastberg, D. & Jocelyn, L. (2005) Achievement and affect in OECD nations. *Oxford Review of Education*. Vol. 31 Issue 4, p517-545.
- Wiseman, A. W. (2012) The Impact of Student Poverty on Science Teaching and Learning: A Cross-National Comparison of the South African Case. *American Behavioral Scientist*, Vol. 56 Issue 7, p941-960.
- Woessmann, L. (2011) Cross-country evidence on teacher performance pay. *Economics of Education Review*. Vol. 30 Issue 3, p404-418.
- Woods-McConney, A., Oliver, M. C., McConney, A., Schibeci, R. & Maor, D. (2014) Science Engagement and Literacy: A retrospective analysis for students in Canada and Australia. *International Journal of Science Education*, Vol. 36 Issue 10, p1588-1608.
- Wu, A. D. & Ercikan, K. (2006) Using Multiple-Variable Matching to Identify Cultural Sources of Differential Item Functioning. *International Journal of Testing*. Vol. 6 Issue 3, p287-300.
- Xin, T., Xu, Z. & Tatsuoaka, K. (2004) Linkage between teacher quality, student achievement, and cognitive skills: A rule-space model. *Studies in Educational Evaluation*. Vol. 30 Issue 4, p205-223.
- Xu, J. (2008) Sibship Size and Educational Achievement: The Role of Welfare Regimes Cross-Nationally. *Comparative Education Review*, Vol. 52 Issue 3, p413-436.
- Yu, C. H. (2012) Examining the relationships among academic self-concept, instrumental motivation, and TIMSS 2007 science scores: a cross-cultural comparison of five East Asian countries/regions and the United States. *Educational Research & Evaluation*. Vol. 18 Issue 8, p713-731.

APPENDIX I

Classification of PISA, TIMSS and CIVED/ICCS articles

This is a revised scheme for classification of the PISA, TIMSS and CIVED/ICCS articles perceived through the Discovery search. The classification has been done within several steps including reading abstracts, but also in most cases the entire articles. The articles have been classified and re-classified several times in different schemes. This means that classification have been managed in several steps (for better descriptions see chapter 5-7 on how these steps are managed).

The first digits in the number stands for which Discovery search the article have been collected; 1 indicates PISA. The different steps in this Discovery search is more in detail explained and discussed in chapter 5; 2 indicate TIMSS, illuminated in chapter 6; 3 indicate CIVED, illuminated in chapter 7, and finally; 4 indicate ICCS also illuminated in chapter 7. Some of the articles are perceived in more than one search. In these cases the articles have been double booked with different individual numbers.

What is presented below is a revised scheme that served as the basis for our coding, mapping, analyses and synthesizes. In the second column from left abbreviations are used for visualizing the classifications. The abbreviations means as follows:

A: stands for articles using the specific ILSA data for investigating different research questions

A1: stands for articles using the specific ILSA data for comparing two or more countries (these articles have later on been coded, mapped, analyzed and synthesized)

A2: stands for articles just discussing results within one single country

B: stands for articles discussing the specific ILSA in policy matters

C: stands for articles discussing methodological matters in relation to the specific ILSA

N: stands for a combined classification of newsletter, editorial texts and other informative texts concerning the specific ILSA

NA: stands for articles not available in full text within our databases. Some of these articles have been classified on abstract level

D: stands for duplicate (these are not presented, just excluded)

M: stands for misplaced (based on content of abstracts, and in some cases of full text articles)

Some of the articles have been classified as Newsletters previously in our presentation. In this appendix some of them are still categorized as newsletters, but when these are not available for us in full text they have been categorized as NA.

Table 1. List of classified articles

| # | C | Reference |
|-------|----|--|
| 1001A | A2 | Mahuteau, S.; Mavromaras, K (2014) An analysis of the impact of socio-economic disadvantage and school quality on the probability of school dropout. <i>Education Economics</i> . Vol.22, No.4, pp.389-411 |
| 1002 | N | Launch of first OECD PISA report on financial literacy. (2014) <i>Education Journal</i> . No.206, p.18 |
| 1003A | A1 | Woods-McConney, A, Oliver, M. C, McConney, A, Schibeci, R. & Maor, D. (2014) Science Engagement and Literacy: A retrospective analysis for students in Canada and Australia. <i>International Journal of Science Education</i> . Vol.36, No. 10 pp.1588-1608. |
| 1004A | A2 | Dumay, X & Dupriez, V (2014) Educational quasi-markets, school effectiveness and social inequalities. <i>Journal of Education Policy</i> . Vol.29, No.4, pp.510-531. |
| 1005 | N | Reid, K (2014) Schleicher throws Wales a lifeline - but huge challenges remain. <i>Education Journal</i> . No.203, p.15. |
| 1006A | A2 | Samuel, R (2014) The gendered interplay between success and well-being during transitions. <i>Educational Research</i> . Vol.56, No.2, pp.202-219. |
| 1007A | A1 | Song, S, Perry, L B & McConney, A (2014) Explaining the achievement gap between Indigenous and non-Indigenous students: an analysis of PISA 2009 results for Australia and New Zealand. <i>Educational Research & Evaluation</i> . Vol.20, No.3, pp.178-198. |
| 1008B | B | Hanberger, A (2014) What PISA Intends To and Can Possibly Achieve: a critical programme theory analysis. <i>European Educational Research Journal</i> . Vol.13, No.2, pp.(missing) |
| 1009A | A1 | Le Donné, N (2014) European Variations in Socioeconomic Inequalities in Students' Cognitive |

| # | C | Reference |
|-------|----|---|
| | | Achievement: The Role of Educational Policies. <i>European Sociological Review</i> . Vol.30, No.3, pp.329-343. |
| 1010 | N | Waterman, C (2014) The Learning Curve keeps on curving. <i>Education Journal</i> . No.198, pp.11-12. |
| 1011 | N | Expansion of numeracy hubs. (2014) <i>Education Journal</i> . No.197, p.8. |
| 1012 | N | PISA: English students perform well in problem solving. (2014) <i>Education Journal</i> . No.197, p.19. |
| 1013B | B | Tveit, S (2014) Educational assessment in Norway. <i>Assessment in Education: Principles, Policy & Practice</i> . Vol.21, No.2, pp.221-237. |
| 1014B | B | Morgan, C & Shahjahan, R A (2014) The legitimization of OECD's global educational governance: examining PISA and AHELO test production. <i>Comparative Education</i> . Vol.50, No.2, pp.192-205. |
| 1015A | A1 | Lee, B (2014) The influence of school tracking systems on educational expectations: a comparative study of Austria and Italy. <i>Comparative Education</i> . Vol.50, No.2, pp.206-228. |
| 1016C | C | Soh, K (2014) Finland and Singapore in PISA 2009: similarities and differences in achievements and school management. <i>Compare: A Journal of Comparative & International Education</i> . Vol.44, No.3, pp.455-471. |
| 1017A | A1 | Jerrim, J & Choi, A (2014) The mathematics skills of school children: how does England compare to the high-performing East Asian jurisdictions? <i>Journal of Education Policy</i> . Vol.29, No.3, pp.349-376. |
| 1018A | A1 | Liem, G A D, Martin, A J, Anderson, M, Gibson, R & Sudmalis, D (2014) The Role of Arts-Related Information and Communication Technology Use in Problem Solving and Achievement: Findings From the Programme for International Student Assessment. <i>Journal of Educational Psychology</i> . Vol.106, No.2, pp.348-363. |
| 1019C | C | Rutkowski, L (2014) Sensitivity of Achievement Estimation to Conditioning Model Misclassification. <i>Applied Measurement in Education</i> . Vol.27, No.2, pp.115-132. |
| 1020C | C | Lenkeit, J & Caro D H (2014) Performance status and change – measuring education system effectiveness with data from PISA 2000–2009. <i>Educational Research & Evaluation</i> . Vol.20, No.2, pp.146-174. |
| 1021A | A2 | Drummond, A & Sauer, J D (2014) Video-Games Do Not Negatively Impact Adolescent Academic Performance in Science, Mathematics or Reading. <i>PLoS ONE</i> . Vol.9, No.4, pp.1-5. |
| 1022 | N | Launch of OECD PISA report on problem solving. (2014) <i>Education Journal</i> . No.196, p.6. |
| 1023 | N | Listening to Andreas. (2014) <i>Education Journal</i> . No.193, p.3. |
| 1024 | N | Waterman, C (2014) It ain't what you know, it's the way that you use it. <i>Education Journal</i> . No.193, pp.21-22. |
| 1025B | B | Gorur, R (2014) Towards a Sociology of Measurement in Education Policy. <i>European Educational Research Journal</i> . Vol.13, No.1, pp.58-72. |
| 1026 | N | Schleicher, A (2014) Education in Japan: Learning to change. <i>Observer</i> . No.298, pp.26-27. |
| 1027 | N | In praise of state education. (2014) <i>Education Journal</i> . No.189, p.3. |
| 1028 | N | PISA finding thrown into doubt? (2014) <i>Education Journal</i> . No.189, p.14. |
| 1029 | D | |
| 1030A | NA | Benito, R, Alegre, M A, Gonzalez-Balletbo, I (2014) School Segregation and Its Effects on Educational Equality and Efficiency in 16 OECD Comprehensive School Systems. <i>Comparative Education Review</i> . Vol.58, No.1, pp.104-134. |
| 1031 | N | The cost of truancy. (2014) <i>Education Journal</i> . No.187, p.17. |
| 1032 | M | Eklöf, H, Pavešič, B J & Grønmo, L S (2014) A Cross-National Comparison of Reported Effort and Mathematics Performance in TIMSS Advanced. <i>Applied Measurement in Education</i> . Vol.27, No.1, pp.31-45. |
| 1033 | N | Improving Pisa scores need library involvement. (2014) <i>CILIP Update</i> . p.10. |
| 1034A | A2 | Belot, M & Vandenberghe, V (2014) Evaluating the 'threat' effects of grade repetition: exploiting the 2001 reform by the French-Speaking Community of Belgium. <i>Education Economics</i> . Vol.22, No.1, pp.73-89. |
| 1035A | A2 | Parker, P D, Marsh, H W, Ciarrochi, J Marshall, S & Abduljabbar, A S (2014) Juxtaposing math self-efficacy and self-concept as predictors of long-term achievement outcomes. <i>Educational Psychology</i> . Vol.34, No.1, pp.29-48. |
| 1036 | N | Morrison, K & van der Werf, G (2014) Editorial. <i>Educational Research & Evaluation</i> . Vol.20, No.1, pp.1-2. |
| 1037A | A1 | Dronkers, J & Kornder, N (2014) Do migrant girls perform better than migrant boys? Deviant gender differences between the reading scores of 15-year-old children of migrants compared to native pupils. <i>Educational Research & Evaluation</i> . Vol.20, No.1, pp.44-66. |
| 1038A | A1 | Olande, O (2014) Graphical artefacts: Taxonomy of students' response to test items. <i>Educational Studies in Mathematics</i> . Vol.85, No.1, pp.53-74. |
| 1039A | A1 | Jerrim, J (2014) The Unrealistic Educational Expectations of High School Pupils: Is America Exceptional? <i>Sociological Quarterly</i> . Vol.55, No.1, pp.196-231. |
| 1040A | NA | Marteletto, L & Andrade, F (2014) The Educational Achievement of Brazilian Adolescents: Cultural Capital and the Interaction between Families and Schools. <i>Sociology of Education</i> . Vol.87, No.1, pp.16-35. |
| 1041 | N | The lucky child. (2013) <i>Education Journal</i> . No.185, p.3. |
| 1042 | N | UK holds steady in latest PISA tests. (2013) <i>Education Journal</i> . No.184, p.1. |
| 1043 | N | Why PISA matters. (2013) <i>Education Journal</i> . No.184, p.3. |

| # | C | Reference |
|-------|----|---|
| 1044 | N | What does the latest PISA really mean? (2013) <i>Education Journal</i> .No.184, pp.4-7. |
| 1045 | N | Richards, C (2013) A sceptical aside. <i>Education Journal</i> . No.184, p.7. |
| 1046 | N | Schleicher, A (2013) The man of the moment. <i>Education Journal</i> . Vol.184, p.8. |
| 1047 | N | Around the world with PISA. (2013) <i>Education Journal</i> . No.184, p.9. |
| 1048 | N | Role of the family crucial in improving performance in schools. (2013) <i>Education Journal</i> . No.184, p.10. |
| 1049 | N | Welsh Government statement on PISA. (2013) <i>Education Journal</i> . No.184, p.11. |
| 1050 | N | Highlights from Scotland's PISA Results. (2013) <i>Education Journal</i> . No.184, p.13. |
| 1051 | N | Taylor, C (2013) What does the OECD's PISA tell us about basic skills? <i>Education Journal</i> . No.184, p.19. |
| 1052 | N | Waterman, C (2013) Michael Gove's "bad policy award" for destabilising teacher training. <i>Education Journal</i> . No.184, pp.22-23. |
| 1053 | N | PISA statement in the Commons. (2013) <i>Education Journal</i> . No.184, pp.27-29. |
| 1054 | N | PISA statement in the Lords. (2013) <i>Education Journal</i> . No.184, pp.29-30. |
| 1055A | A2 | Lounkaew, K (2013) Explaining urban-rural differences in educational achievement in Thailand: Evidence from PISA literacy data. <i>Economics of Education Review</i> . Vol.37, pp.213-225. |
| 1056A | A2 | Kiss, D (2013) Are immigrants and girls graded worse? Results of a matching approach. <i>Education Economics</i> Vol.21, No.5, pp.447-463. |
| 1057 | N | All students have the potential to become high achievers in the right conditions. (2013) <i>Education Journal</i> . No.181, p.3. |
| 1058 | N | Coryton, L (2013) All students have the potential to become high achievers in the right conditions, argued PISA. <i>Education Journal</i> . No.181, p.11. |
| 1059 | N | Reid, K (2013) PISA 2009: The turbulent aftermath in Wales - Part 2. <i>Education Journal</i> . No.181, pp.12-17. |
| 1060 | N | The judgement of PISA cometh. (2013) <i>Education Journal</i> . No.180, p.3. |
| 1061 | N | OECD to launch PISA test for schools in England. (2013) <i>Education Journal</i> . No.180, p.4. |
| 1062 | N | Reform and muddle. (2013) <i>Education Journal</i> . No.179, p.3. |
| 1063 | N | Schweisfurth, M (2013) Perennial issues, contemporary variants: comparative education research as a mirror to the social world. <i>Comparative Education</i> . Vol.49, No.4, pp.405-406. |
| 1064B | B | Sellar, S & Lingard, B (2013) Looking East: Shanghai, PISA 2009 and the reconstitution of reference societies in the global education policy field. <i>Comparative Education</i> . Vol.49, No.4, pp.464-485. |
| 1065 | N | France and PISA: Spurring reform? (2013) <i>OECD Observer</i> . No.297, p.19. |
| 1066 | N | Class progress. (2013) <i>OECD Observer</i> . No.297, p.20. |
| 1067 | N | Some country snapshots. (2013) <i>OECD Observer</i> . No.297, p.21. |
| 1068 | N | Overall trends. (2013) <i>OECD Observer</i> . No.297, p.21. |
| 1069 | N | Schleicher, A (2013) PISA's hard lessons. <i>OECD Observer</i> . No.297, p.22. |
| 1070 | N | Venturi, R (2013) Creative minds. <i>OECD Observer</i> . No.297, pp.23-24. |
| 1071 | N | Which car? A test from PISA 2012. (2013) <i>OECD Observer</i> . No.297, pp.24-26. |
| 1072 | N | Marginson, S (2013) Foreword. <i>Oxford Studies in Comparative Education</i> . Vol.23, No.2, pp.7-9. |
| 1073B | B | Baxter, J & Clark, J (2013) Farewell to the tick box inspector? Ofsted and the changing regime of school inspection in England. <i>Oxford Review of Education</i> . Vol.39, No.5, pp.702-718. |
| 1074B | B | Martino, W & Rezai-Rashti, G (2013) 'Gap talk' and the global rescaling of educational accountability in Canada. <i>Journal of Education Policy</i> . Vol.28, No.5, pp.589-611. |
| 1075B | B | Sellar, S & Lingard, B (2013) The OECD and global governance in education. <i>Journal of Education Policy</i> . Vol.28, No.5, pp.710-725. |
| 1076B | B | Kind, P M (2013) Conceptualizing the Science Curriculum: 40 Years of Developing Assessment Frameworks in Three Large-Scale Assessments. <i>Science Education</i> . Vol.97, No.5, pp.671-694. |
| 1077A | A2 | Segeritz, M & Pant, H A (2013) Do They Feel the Same Way About Math?: Testing Measurement Invariance of the PISA "Students' Approaches to Learning" Instrument Across Immigrant Groups Within Germany. <i>Educational & Psychological Measurement</i> . Vol.73, No.4, pp.601-630. |
| 1078A | A1 | Li, H, Lei, P-W & Pace, C L (2013) Reading subskill differences between students in Shanghai-China and the US: evidence from PISA 2009. <i>Educational Research & Evaluation</i> . Vol.19, No.6, pp.490-509. |
| 1079A | A2 | Palidano, C, Hanel, B & Buddelmeyer, H (2013) Explaining the socio-economic status school completion gap. <i>Education Economics</i> . Vol.21, No.3, pp.230-247. |
| 1080A | A2 | Woolard, J C (2013) Prelude to the Common Core: Internationally Benchmarking a State's Math Standards. <i>Educational Policy</i> . Vol.27, No.4, pp.615-644. |
| 1081A | A1 | Merry, J J (2013) Tracing the U.S. Deficit in PISA Reading Skills to Early Childhood: Evidence from the United States and Canada. <i>Sociology of Education</i> . Vol.86, No.3, pp.234-252. |
| 1082B | B | Machin, S, McNally, S & Wyness, G (2013) Educational attainment across the UK nations: performance, inequality and evidence. <i>Educational Research</i> . Vol.55, No.2, pp.139-164. |
| 1083A | A2 | Hampden-Thompson, G & Bennett, J (2013) Science Teaching and Learning Activities and Students' Engagement in Science. <i>International Journal of Science Education</i> . Vol.35, No.8, pp.1325-1343. |
| 1084 | N | PISA: Rasch Model. (2013) <i>Education Journal</i> . No.165, p.15. |
| 1085 | D | |

| # | C | Reference |
|-------|----|---|
| 1086A | A2 | Felouiz, G & Charmillot, S (2013) School tracking and educational inequality: a comparison of 12 education systems in Switzerland. <i>Comparative Education</i> . Vol.49, No.2, pp.181-205. |
| 1087 | M | Andrews, P (2013) Finnish Mathematics Teaching from a Reform Perspective: A Video-Based Case-Study Analysis. <i>Comparative Education Review</i> . Vol.57, No.2, pp.189-211. |
| 1088A | A2 | Safarzyńska, K (2013) Socio-economic Determinants of Demand for Private Tutoring. <i>European Sociological Review</i> . Vol.29, No.2, pp.139-154. |
| 1089B | B | Meyer, H-D & Benavot, A (2013) PISA and the Globalization of Education Governance: some puzzles and problems. <i>Oxford Studies in Comparative Education</i> . Vol.23, No.1, pp.9-26. |
| 1090B | B | Owens, T L (2013) Thinking Beyond League Tables: a review of key PISA research questions. <i>Oxford Studies in Comparative Education</i> . Vol.23, No.1, pp.27-49. |
| 1091B | B | Varjo, J, Simola, H & Rinne, R (2013) Finland's PISA Results: an analysis of dynamics in education politics. <i>Oxford Studies in Comparative Education</i> . Vol.23, No.1, pp.51-76. |
| 1092A | NA | Andrews, P (2013) What Does PISA Performance Tell Us about Mathematics Teaching Quality? Case Studies from Finland and Flanders. <i>Oxford Studies in Comparative Education</i> . Vol.23, No.1, pp.99-115. |
| 1093B | B | Kamens, D H (2013) Globalization and the Emergence of an Audit Culture: PISA and the search for 'best practices' and magic bullets. <i>Oxford Studies in Comparative Education</i> . Vol.23, No.1, pp.117-139. |
| 1094B | B | Tröhler, D (2013) The OECD and Cold War Culture: thinking historically about PISA. <i>Oxford Studies in Comparative Education</i> . Vol.23, No.1, pp.141-161. |
| 1095B | B | Lockheed, M (2013) Causes and Consequences of International Assessments in Developing Countries. <i>Oxford Studies in Comparative Education</i> . Vol.23, No.1, pp.163-183. |
| 1096B | B | Sellar, S & Lingard, B (2013) PISA and the Expanding Role of the OECD in Global Educational Governance. <i>Oxford Studies in Comparative Education</i> . Vol.23, No.1, pp.185-206. |
| 1097B | B | Meyer, H-D & Schiller, K (2013) Gauging the Role of Non-educational Effects in Large-scale Assessments: socio-economics, culture and PISA outcomes. <i>Oxford Studies in Comparative Education</i> . Vol.23, No.1, pp.207-224. |
| 1098B | B | Yong, Z & Meyer, H-D (2013) High on PISA, Low on Entrepreneurship? What PISA Does Not Measure. <i>Oxford Studies in Comparative Education</i> . Vol.23, No.1, pp.267-278. |
| 1099B | B | Wiseman, A W (2013) Policy Responses to PISA in Comparative Perspective. <i>Oxford Studies in Comparative Education</i> . Vol.21, No.1, pp.303-322. |
| 1100 | N | Link between reading levels and social disadvantage.(2013) <i>CILIP Update</i> . Vol.12, No.3, p.17. |
| 1101B | B | Anagnostopoulou, K, Hatzinikita, V, Christidou, V & Dimopoulos, K (2013) PISA Test Items and School-Based Examinations in Greece: Exploring the relationship between global and local assessment discourses. <i>International Journal of Science Education</i> . Vol.35, No.4, pp.636-662. |
| 1102 | N | Video-watch: Education performers and reformers. (2013) <i>OECD Observer</i> . No.294, p.36. |
| 1103A | A1 | Stoet, G & Geary, D C (2013) Sex Differences in Mathematics and Reading Achievement Are Inversely Related: Within- and Across-Nation Assessment of 10 Years of PISA Data. <i>PLoS ONE</i> . Vol.8, No.3, pp.1-10. |
| 1104A | A2 | Samuel, R, Bergman, M & Hupka-Brunner, S (2013) The Interplay between Educational Achievement, Occupational Success, and Well-Being. <i>Social Indicators Research</i> . Vol.111, No.1, pp.75-96. |
| 1105 | N | Going global. (2013) <i>Education Journal</i> . No.153, p.2. |
| 1106 | N | Key, S (2013) PISA in Focus supports equity and quality claim. <i>Education Journal</i> . No.153, p.9. |
| 1107A | NA | Goos, M, Schreier, B M, Knipprath, H M D, De Fraine, B, Van Damme, J & Trautwein, U (2013) How Can Cross-Country Differences in the Practice of Grade Retention Be Explained? A Closer Look at National Educational Policy Factors. <i>Comparative Education Review</i> . Vol.57, No.1, pp.54-84. |
| 1108 | M | Mäkinen, M (2013) Towards community oriented curriculum in Finnish literacy education. <i>Journal of Teacher Education</i> . Vol.36, No.1, pp.97-112. |
| 1109C | C | Täht, K & Must, O (2013) Comparability of educational achievement and learning attitudes across nations. <i>Educational Research & Evaluation</i> . Vol.19, No.1, pp.19-38. |
| 1110 | D | |
| 1111C | C | Micklewright, J, Schnepf, S V & Silva, P N (2012) Peer effects and measurement error: The impact of sampling variation in school survey data (evidence from PISA). <i>Economics of Education Review</i> . Vol.31, No.6, pp.1136-1142. |
| 1112 | N | Does the maths add up? (2012) <i>Education Journal</i> . No.145, p.2. |
| 1113 | N | Baird, J-A (2012) Editorial. <i>Assessment in Education: Principles, Policy & Practice</i> . Vol.19, No.4, pp.389-391. |
| 1114B | B | Dominguez, M, Vieira, M-J & Vidal, J (2012) The impact of the Programme for International Student Assessment on academic journals. <i>Assessment in Education: Principles, Policy & Practice</i> . Vol.19, No.4, pp.393-409. |
| 1115 | N | Morrison, K & van der Werf, G (2012) Editorial. <i>Educational Research & Evaluation</i> . Vol.18, No.8, pp.691-692. |
| 1116A | A1 | Sikora, J & Pokropek, A (2012) Intergenerational Transfers of Preferences for Science Careers in Comparative Perspective. <i>International Journal of Science Education</i> . Vol.34, No.16, pp.2501-2527. |

| # | C | Reference |
|-------|----|---|
| 1117A | NA | Nagengast, B & Marsh, H W (2012) Big Fish in Little Ponds Aspire More: Mediation and Cross-Cultural Generalizability of School-Average Ability Effects on Self-Concept and Career Aspirations in Science. <i>Journal of Educational Psychology</i> . Vol.104, No.4, pp.1033-1053. |
| 1118A | NA | Martin, A J, Liem, G A D, Mok, M M C & Xu, J (2012) Problem Solving and Immigrant Student Mathematics and Science Achievement: Multination Findings from the Programme for International Student Assessment (PISA). <i>Journal of Educational Psychology</i> . Vol.104, No.4, pp.1054-1073. |
| 1119 | D | |
| 1120 | D | |
| 1121A | A1 | Gamboa, L F & Waltenberg, F D (2012) Inequality of opportunity for educational achievement in Latin America: Evidence from PISA 2006–2009. <i>Economics of Education Review</i> . Vol.31, No.5, pp.694-708. |
| 1122C | C | Huang, C (2012) Discriminant and incremental validity of self-concept and academic self-efficacy: a meta-analysis. <i>Educational Psychology</i> . Vol.32, No.6, pp.777-805. |
| 1123A | A2 | Sun, L, Bradley, K D & Akers, K (2012) A Multilevel Modelling Approach to Investigating Factors Impacting Science Achievement for Secondary School Students: PISA Hong Kong Sample. <i>International Journal of Science Education</i> . Vol.34, No.14, pp.2107-2125. |
| 1124 | N | OECD PISA study: Improving science. (2012) <i>Education Journal</i> . No.136, p.4. |
| 1125B | B | Feniger, Y, Livneh, I & Yogev, A (2012) Globalisation and the politics of international tests: the case of Israel. <i>Comparative Education</i> . Vol.48, No.3, pp.323-335. |
| 1126A | A1 | Rutkowski, L, Rutkowski, D & Plucker, J (2012) International determinants of private school attendance. <i>Educational Research & Evaluation</i> . Vol.18, No.6, pp.375-391. |
| 1127 | N | Smyth, G (2012) Untitled. <i>European Journal of Teacher Education</i> . Vol.35, No.3, pp.255-257. |
| 1128B | B | Chung, J, Atkin, C & Moore, J (2012) The rise and fall of the MTL: an example of European policy borrowing. <i>European Journal of Teacher Education</i> . Vol.35, No.3, pp.259-274. |
| 1129 | N | Learning from Finland. (2012) <i>Education Journal</i> . No.135, p.2. |
| 1130B | B | Gaber, S, Cankar, G, Umek L M & Tasner, V (2012) The danger of inadequate conceptualisation in PISA for education policy. <i>Compare: A Journal of Comparative & International Education</i> . Vol.42, No.4, pp.647-663. |
| 1131C | C | Oliveri, M E, Olson, B F, Ercikan, K & Zumbo, B D (2012) Methodologies for Investigating Item- and Test-Level Measurement Equivalence in International Large-Scale Assessments. <i>International Journal of Testing</i> . Vol.12, No.3, pp.203-223. |
| 1132A | A1 | Reilly, D (2012) Gender, Culture, and Sex-Typed Cognitive Abilities. <i>PLoS ONE</i> . Vol.7, No.7, pp.1-16. |
| 1133A | A2 | Byun, S, Schofer, E & Kim, K (2012) Revisiting the Role of Cultural Capital in East Asian Educational Systems: The Case of South Korea. <i>Sociology of Education</i> . Vol.85, No.3, pp.219-239. |
| 1134 | N | Rewarding Pay. (2012) <i>Education Journal</i> . No.131, p.9. |
| 1135 | N | Howson, J (2012) Can't wait for PISA. <i>Education Journal</i> . No.131, p.14. |
| 1136 | N | CESE Publications. (2012) <i>European Education</i> . Vol.44, No.1, pp.104-105. |
| 1137A | A1 | Sikora, J & Pokropek, A (2012) Gender Segregation of Adolescent Science Career Plans in 50 Countries. <i>Science Education</i> . Vol.96, No.2, pp.234-264. |
| 1138A | NA | Jun, X & Hampden-Thompson, G (2012) Cultural Reproduction, Cultural Mobility, Cultural Resources, or Trivial Effect? A Comparative Approach to Cultural Capital and Educational Performance. <i>Comparative Education Review</i> . Vol.56, No.1, pp.98-124. |
| 1139 | D | |
| 1140A | NA | Dalton, B (2012) Grade Level and Science Achievement: US Performance in Cross-National Perspective. <i>Comparative Education Review</i> . Vol.56, No.1, pp.125-154. |
| 1141A | A2 | Yildirim, S (2012) Teacher Support, Motivation, Learning Strategy Use, and Achievement: A Multilevel Mediation Model. <i>Journal of Experimental Education</i> . Vol.80, No.2, pp.150-172. |
| 1142 | D | |
| 1143A | A1 | Shapira, M (2012) An Exploration of Differences in Mathematics Attainment among Immigrant Pupils in 18 OECD Countries. <i>European Educational Research Journal</i> . Vol.11, No.1, pp.68-95. |
| 1144B | B | Pons, X (2012) Going beyond the "PISA Shock" Discourse: An Analysis of the Cognitive Reception of PISA in Six European Countries, 2001-2008. <i>European Educational Research Journal</i> . Vol.11, No.2, pp.206-226. |
| 1145B | B | Carvalho, L M (2012) The Fabrications and Travels of a Knowledge-Policy Instrument. <i>European Educational Research Journal</i> . Vol.11, No.2, pp.172-188. |
| 1146B | B | Mangez, E & Hilgers, M (2012) The Field of Knowledge and the Policy Field in Education: PISA and the Production of Knowledge for Policy. <i>European Educational Research Journal</i> . Vol.11, No.2, pp.189-205. |
| 1147B | B | Neumann, E, Kiss, A & Feles, I (2012) The Hard Work of Interpretation: The National Politics of PISA Reception in Hungary and Romania. <i>European Educational Research Journal</i> . Vol.11, No.2, pp.227-242. |
| 1148B | B | Grek, S (2012) What PISA Knows and Can Do: Studying the Role of National Actors in the Making of PISA. <i>European Educational Research Journal</i> . Vol.11, No.2, pp.243-254. |
| 1149A | A1 | Dronkers, J, van der Velden, R & Dunne, A (2012) Why Are Migrant Students Better off in Certain Types of Educational Systems or Schools than in Others? <i>European Educational Research Journal</i> . Vol.11, No.1, |

| # | C | Reference |
|-------|----|--|
| | | pp.11-44. |
| 1150 | M | Moran, A & Clarke, L (2012) Back to the Future: Do Lessons from Finland Point the Way to a Return to Model Schools for Northern Ireland? <i>European Journal of Teacher Education</i> . Vol.35, No.3, pp.275-288. |
| 1151B | B | Gür, B S, Celik, Z & Özoglu, M (2012) Policy options for Turkey: a critique of the interpretation and utilization of PISA results in Turkey. <i>Journal of Education Policy</i> . Vol.27, No.1, pp.1-21. |
| 1152B | B | Dobbins, M & Martens, K (2012) Towards an education approach à la finlandaise? French education policy after PISA. <i>Journal of Education Policy</i> . Vol.27, No.1, pp.23-43. |
| 1153B | B | Hartong, S (2012) Overcoming Resistance to Change: PISA, School Reform in Germany and the Example of Lower Saxony. <i>Journal of Education Policy</i> . Vol.27, No.6, pp.747-760. |
| 1154 | D | |
| 1155 | D | |
| 1156 | D | |
| 1157 | M | Pereyra, M & Kotthoff, H-G (n.d) CESE News |
| 1158 | N | Looney, J & Gordon, J (2011) Editorial. <i>European Journal of Education</i> . Vol.46, No.4, pp.433-439. |
| 1159B | B | Pons, X (2011) What Do We Really Learn from PISA? The Sociology of its Reception in Three European Countries (2001-2008). <i>European Journal of Education</i> . Vol.46, No.4, pp.540-548. |
| 1160B | B | Lingard, B & Rawolle, S (2011) New scalar politics: implications for education policy. <i>Comparative Education</i> . Vol.47, No.4, pp.489-502. |
| 1161 | N | George F. Bereday Award. (2011) <i>Comparative Education Review</i> . Vol.55, No.4, p.634. |
| 1162 | N | Taking stock of skills. (2011) <i>OECD Observer</i> . No.287, pp.21-22. |
| 1163C | C | Oliveri, M E & Ercikan, K (2011) Do Different Approaches to Examining Construct Comparability in Multilanguage Assessments Lead to Similar Conclusions? <i>Applied Measurement in Education</i> . Vol.24, No.4, pp.349-366. |
| 1164A | A2 | McConney, A, Oliver, M, Woods-McConney, A & Schibeci, R (2011) Bridging the Gap? A Comparative, Retrospective Analysis of Science Literacy and Interest in Science for Indigenous and Non-Indigenous Australian Students. <i>International Journal of Science Education</i> . Vol.33, No.14, pp.2017-2035. |
| 1165A | A2 | Tshering, G & Prain, V (2011) Benchmarking the performance of Bhutanese students with the performance of the students from the OECD's PISA countries. <i>Educational Research & Evaluation</i> . Vol.17, No.4, pp.263-281. |
| 1166A | A2 | Liu, X & Whitford, M (2011) Opportunities-to-Learn at Home: Profiles of Students with and without Reaching Science Proficiency. <i>Journal of Science Education and Technology</i> . Vol.20, No.4, pp.375-387. |
| 1167A | A1 | Bratti, M, Checchi, D & Filippin, A (2011) Should you compete or cooperate with your schoolmates? <i>Education Economics</i> . Vol.19, No.3, pp.275-289. |
| 1168A | A1 | Woessmann, L (2011) Cross-country evidence on teacher performance pay. <i>Economics of Education Review</i> . Vol.30, No.3, pp.404-418. |
| 1169 | N | Around the world. (2011) <i>Education Journal</i> . No.127, p.6. |
| 1170 | N | Mansell, W (2011) What is it about Finland? <i>Education Journal</i> . No.127, pp.24-25. |
| 1171 | N | Reid, K (2011) PISA and all that! <i>Education Journal</i> . No.127, p.27. |
| 1172 | N | Husbands, C (2011) Research and policy. <i>Education Journal</i> . No.127, p.32. |
| 1173 | N | Welsh Assembly - PISA and Funding. (2011) <i>Education Journal</i> . No.127, p.42. |
| 1174C | C | Frey, A & Seitz N-N (2011) Hypothetical Use of Multidimensional Adaptive Testing for the Assessment of Student Achievement in the Programme for International Student Assessment. <i>Educational & Psychological Measurement</i> . Vol.71, No.3, pp.503-522. |
| 1175A | A2 | Green, A (2011) Lifelong Learning, Equality and Social Cohesion. <i>European Journal of Education</i> . Vol.46, No.2, pp.228-243. |
| 1176 | N | Education: Bridging the classroom divide. (2011) <i>OECD Observer</i> . No.284, pp.50-51. |
| 1177C | C | Hopfenbeck, T N & Maul, A (2011) Examining Evidence for the Validity of PISA Learning Strategy Scales Based on Student Response Processes. <i>International Journal of Testing</i> . Vol.11, No.2, pp.95-121. |
| 1178B | B | Bieber, T & Martens, K (2011) The OECD PISA Study as a Soft Power in Education? Lessons from Switzerland and the US. <i>European Journal of Education</i> . Vol.46, No.1, pp.101-116. |
| 1179 | N | Top of the class. (2011) <i>OECD Observer</i> . No.282/283, p.44. |
| 1180A | A2 | Meunier, M (2011) Immigration and student achievement: Evidence from Switzerland. <i>Economics of Education Review</i> . Vol.30, No.1, pp.16-38. |
| 1181A | A2 | Perelman, S & Santin, D (2011) Measuring educational efficiency at student level with parametric stochastic distance functions: an application to Spanish PISA results. <i>Education Economics</i> . Vol.19, No.1, pp.29-49. |
| 1182B | B | Gorur, R (2011) Policy as Assemblage. <i>European Educational Research Journal</i> . Vol.10, No.4, pp.611-622. |
| 1183A | A1 | Bybee, R & McCrae, B (2011) Scientific Literacy and Student Attitudes: Perspectives from PISA 2006 science. <i>International Journal of Science Education</i> . Vol.33, No.1, pp.7-26. |
| 1184C | C | Krapp, A & Prenzel, M (2011) Research on Interest in Science: Theories, methods, and findings. <i>International Journal of Science Education</i> . Vol.33, No.1, pp.27-50. |

| # | C | Reference |
|-------|----|---|
| 1185A | A1 | Ainley, M & Ainley, J (2011) A Cultural Perspective on the Structure of Student Interest in Science. <i>International Journal of Science Education</i> . Vol.33, No.1, pp.51-71. |
| 1186C | C | Drechsel, B, carstensen, C & Prenzel, M (2011) The Role of Content and Context in PISA Interest Scales: A study of the embedded interest items in the PISA 2006 science assessment. <i>International Journal of Science Education</i> . Vol.33, No.1, pp.73-95. |
| 1187A | A1 | Olsen, R V & Lie, S (2011) Profiles of Students' Interest in Science Issues around the World: Analysis of data from PISA 2006. <i>International Journal of Science Education</i> . Vol.33, No.1, pp.97-120. |
| 1188A | A1 | Kjærnsli, M & Lie, S (2011) Students' Preference for Science Careers: International comparisons based on PISA 2006. <i>International Journal of Science Education</i> . Vol.33, No.1, pp.121-144. |
| 1189A | A1 | Basl, J (2011) Effect of School on Interest in Natural Sciences: A comparison of the Czech Republic, Germany, Finland, and Norway based on PISA 2006. <i>International Journal of Science Education</i> . Vol.33, No.1, pp.145-157. |
| 1190A | A1 | Buccheri, G, Gurber, N A & Bruhwiler, C (2011) The Impact of Gender on Interest in Science Topics and the Choice of Scientific and Technical Vocations. <i>International Journal of Science Education</i> . Vol.33, No.1, pp.159-178. |
| 1191 | D | |
| 1192A | A1 | Montt, G (2011) Cross-national Differences in Educational Achievement Inequality. <i>Sociology of Education</i> . Vol.84, No.1, pp.49-68. |
| 1193B | B | Alexander, R J (2010) 'World class schools' - noble aspiration or globalised hokum? <i>Compare: A Journal of Comparative & International Education</i> . Vol.40, No.6, pp.801-817. |
| 1194A | A1 | Martins, L & Veiga, P (2010) Do inequalities in parents' education play an important role in PISA students' mathematics achievement test score disparities? <i>Economics of Education Review</i> . Vol.29, No.6, pp.1016-1033. |
| 1195 | N | Everybody loves PISA. (2010) <i>Education Journal</i> . No.126, p.6. |
| 1196 | N | Bray, N (2010) PISA. <i>Education Journal</i> . No.126, pp.30-31. |
| 1197A | A1 | Mostafa, T (2010) Decomposing inequalities in performance scores: the role of student background, peer effects and school characteristics. <i>International Review of Education/Internationale Zeitschrift für Erziehungswissenschaft</i> . Vol.56, No.5/6, pp.567-589. |
| 1198B | B | Grek, S (2010) International Organisations and the Shared Construction of Policy 'Problems': problematisation and change in education governance in Europe. <i>European Educational Research Journal</i> . Vol.9, No.3, pp.7- |
| 1199A | A2 | Marks, G N (2010) What aspects of schooling are important? School effects on tertiary entrance performance. <i>School Effectiveness & School Improvement</i> . Vol.21, No.3, pp.267-287. |
| 1200A | A1 | Mei-Shiu, C (2010) Effects of science interest and environmental responsibility on science aspiration and achievement: gender differences and cultural supports. <i>Educational Research & Evaluation</i> . Vol.16, No.4, pp.345-370. |
| 1201B | B | Cung, J & Tsuruta, Y (2010) An investigation of reasons for Finland's success in PISA. <i>Compare: A Journal of Comparative & International Education</i> . Vol.40, No.4, pp.551-553. |
| 1202A | A1 | Alegre, M A & Ferrer, G (2010) School regimes and education equity: some insights based on PISA 2006. <i>British Educational Research Journal</i> . Vol.36, No.3, pp.433-461. |
| 1203B | B | Polesel, J (2010) Educational outcomes, policy reform and cultural capital in Italian secondary schools. <i>Comparative Education</i> . Vol.46, No.2, pp.173-187. |
| 1204A | A1 | Tramonte, L & Willms, J D (2010) Cultural capital and its effects on education outcomes. <i>Economics of Education Review</i> . Vol.29, No.2, pp.200-213. |
| 1205A | A1 | Nonoyama-Tarumi, Y & Willms, J D (2010) The relative and absolute risks of disadvantaged family background and low levels of school resources on student literacy. <i>Economics of Education Review</i> . Vol.29, No.2, pp.214-224. |
| 1206A | A1 | Robert, P (2010) Social origin, school choice, and student performance. <i>Educational Research & Evaluation</i> . Vol.16, No.2, pp.107-129. |
| 1207B | B | Gauthier, R-F & Le Gouvello, M (2010) The French Curricular Exception and the Troubles of Education and Internationalisation: will it be enough to 'rearrange the deckchairs'? <i>European Journal of Education</i> . Vol.45, No.1, pp.74-88. |
| 1208B | B | Hofman, W H A, Hofman, R H & Gray, J M (2010) Institutional Contexts and International Performances in Schooling: comparing patterns and trends over time in international surveys. <i>European Journal of Education</i> . Vol.45, No.1, pp.153-173. |
| 1209C | C | Chen, S-S & Luoh, M-C (2010) Are Mathematics and Science Test Scores Good Indicators of Labor-Force Quality? <i>Social Indicators Research</i> . Vol.96, No.1, pp.133-143. |
| 1210A | NA | 'Cross-national patterns of gender differences in mathematics: A meta-analysis' Correction. (2010) <i>Psychological Bulletin</i> . Vol.136, No.2. |
| 1211A | NA | Capraro, R M, Capraro, M M, Yerkiner, Z E, Ozel, S, Kim, H G & Küçük, A R (2010) An international comparison of grade 6 students' understanding of the equal sign. <i>Psychological Reports</i> . Vol.106, No.1, pp.49-53. |

| # | C | Reference |
|-------|----|--|
| 1212A | A1 | McDaniel, A (2010) Cross-National Gender Gaps in Educational Expectations: The Influence of National-Level Gender Ideology and Educational Systems. <i>Comparative Education Review</i> . Vol.54, No.1, pp.27-50. |
| 1213B | B | Takayama, K (2010) Politics of Externalization in Reflexive Times: Reinventing Japanese Education Reform Discourses through "Finnish PISA Success". <i>Comparative Education Review</i> . Vol.54, No.1, pp.51-75. |
| 1214 | M | Ostermeier, C, Prenzel, M & Duit, R (2010) Improving Science and Mathematics Instruction: The SINUS Project as an example for reform as teacher professional development. <i>International Journal of Science Education</i> . Vol.32, No.3, pp.303-327. |
| 1215 | M | Watson, A, Kehler, M & Martino, W (2010) The Problem of Boys' Literacy Underachievement: Raising Some Questions. <i>Journal of Adolescent & Adult Literacy</i> . Vol.53, No.5, pp.356-361. |
| 1216B | B | Ringarp, J & Rothland, M (2010) Is the Grass Always Greener? The Effect of the PISA Results on Education Debates in Sweden and Germany. <i>European Educational Research Journal</i> . Vol.9, No.3, pp.422-430. |
| 1217A | A1 | Else-Quest, N M, Hyde, J S & Linn, M C (2010) Cross-National Patterns of Gender Differences in Mathematics: A Meta-Analysis. <i>Psychological Bulletin</i> . Vol.136, No.1, pp.103-127. |
| 1218 | D | |
| 1219 | M | Kramarski, B & Revach, T (2009) The challenge of self-regulated learning in mathematics teachers' professional training. <i>Educational Studies in Mathematics</i> . Vol.72, No.3, pp.379-399. |
| 1220A | A1 | Dettmers, S, Trautwein, U & Ludtke, O (2009) The relationship between homework time and achievement is not universal: evidence from multilevel analyses in 40 countries. <i>School Effectiveness & School Improvement</i> . Vol.20, No.4, pp.375-405. |
| 1221 | M | Pinto, R & El Boudamoussi, S (2009) Scientific Processes in PISA Tests Observed for Science Teachers. <i>International Journal of Science Education</i> . Vol.31, No.16, pp.2137-2159. |
| 1222A | A2 | Escardibul, J-O & Villarroya, A (2009) The inequalities in school choice in Spain in accordance to PISA data. <i>Journal of Education Policy</i> . Vol.24, No.6, pp.673-696. |
| 1223 | D | |
| 1224 | M | Hambleton, R K, Sireci, S G & Smith, Z R (2009) How Do Other Countries Measure Up to the Mathematics Achievement Levels on the National Assessment of Educational Progress? <i>Applied Measurement in Education</i> . Vol.22, No.4, pp.376-393. |
| 1225A | A2 | Notten, N, Peter, J, Kraaykamp, G & Valkenburg, P M (2009) Research Note: Digital Divide Across Borders—A Cross-National Study of Adolescents' Use of Digital Technologies. <i>European Sociological Review</i> . Vol.25, No.5, pp.551-560. |
| 1226B | B | Rautalin, M & Alasuutari, P (2009) The uses of the national PISA results by Finnish officials in central government. <i>Journal of Education Policy</i> . Vol.24, No.5, pp.539-556. |
| 1227A | NA | Horn, D (2009) Age of selection counts: a cross-country analysis of educational institutions. <i>Educational Research & Evaluation</i> . Vol.15, No.4, pp.343-366. |
| 1228A | NA | Notten, N & Kraaykamp, G (2009) Home media and science performance: a cross-national study. <i>Educational Research & Evaluation</i> . Vol.15, No.4, pp.367-384. |
| 1229A | NA | Sikora, J & Saha, L J (2009) Gender and professional career plans of high school students in comparative perspective. <i>Educational Research & Evaluation</i> . Vol.15, No.4, pp.385-403. |
| 1230A | A1 | Hampden-Thompson, G (2009) Are two better than one? A comparative study of achievement gaps and family structure. <i>Compare: A Journal of Comparative & International Education</i> . Vol.39, No.4, pp.513-529. |
| 1231A | NA | Seaton, M, Marsh, H W & Craven, R G (2009) Earning Its Place as a Pan-Human Theory: Universality of the Big-Fish-Little-Pond Effect Across 41 Culturally and Economically Diverse Countries. <i>Journal of Educational Psychology</i> . Vol.101, No.2, pp.403-419. |
| 1232 | D | |
| 1233A | A2 | Liu, O L & Wilson, M (2009) Gender Differences in Large-Scale Math Assessments: PISA Trend 2000 and 2003. <i>Applied Measurement in Education</i> . Vol.22, No.2, pp.164-184. |
| 1234C | C | Hüseyin, H Y & Berberoğlu, G (2009) Judgmental and Statistical DIF Analyses of the PISA-2003 Mathematics Literacy Items. <i>International Journal of Testing</i> . Vol.9, No.2, pp.108-121. |
| 1235A | A1 | Perry, L (2009) Characteristics of Equitable Systems of Education: A Cross-National Analysis. <i>European Education</i> . Vol.41, No.1, pp.79-100. |
| 1236A | A2 | Close, S & Shiel, G (2009) Gender and PISA Mathematics: Irish results in context. <i>European Educational Research Journal</i> . Vol.8, No.1, pp.2- |
| 1237A | NA | Halldórsson, A M & Ólafsson, R F (2009) The Case of Iceland in PISA: girls' educational advantage. <i>European Educational Research Journal</i> . Vol.8, No.1, pp.3- |
| 1238A | A1 | Fredriksson, U, Holzer, T, McCluskey-Cavin, H & Taube, K (2009) Strengths and Weaknesses in the Swedish and Swiss Education Systems: a comparative analysis based on PISA data. <i>European Educational Research Journal</i> . Vol.8, No.1, pp.4- |
| 1239A | A1 | Lafontaine, D & Monseur, C (2009) Gender Gap in Comparative Studies of Reading Comprehension: to what extent do the test characteristics make a difference? <i>European Educational Research Journal</i> . Vol.8, |

| # | C | Reference |
|-------|----|---|
| | | No.1, pp.5- |
| 1240A | A1 | Liu, O L & Wilson, M (2009) Gender Differences and Similarities in PISA 2003 Mathematics: A Comparison between the United States and Hong Kong. <i>International Journal of Testing</i> . Vol.9, No.1, pp.20-40. |
| 1241 | M | Tjeldvoll, A (2008/2009) Finnish Higher Education Reforms: Responding to Globalization. <i>European Education</i> . Vol.40, No.4, pp.93-107. |
| 1242 | D | |
| 1243 | D | |
| 1244B | B | Grek, S (2009) Governing by numbers: the PISA 'effect' in Europe. <i>Journal of Education Policy</i> , Vol.24, No.1, pp.23-37. |
| 1245 | D | |
| 1246B | B | Bybee, R (2008) Scientific Literacy, Environmental Issues, and PISA 2006: The 2008 Paul F-Brandwein Lecture. <i>Journal of Science Education and Technology</i> . Vol.17, No.6, pp.566-585. |
| 1247A | A2 | Linnakylä, P & Malin, A (2008) Finnish Students' School Engagement Profiles in the Light of PISA 2003. <i>Scandinavian Journal of Educational Research</i> . Vol.52, No.6, pp.583-602. |
| 1248 | N | Goldstein, H & Thomas, S M (2008) Reflections on the international comparative surveys debate. <i>Assessment in Education: Principles, Policy & Practice</i> . Vol.15, No.3, pp.215-222. |
| 1249B | B | McGaw, B (2008) The role of the OECD in international comparative studies of achievement. <i>Assessment in Education: Principles, Policy & Practice</i> . Vol.15, No.3, pp.223-243. |
| 1250B | B | Egelund, N (2008) The value of international comparative studies of achievement - a Danish perspective. <i>Assessment in Education: Principles, Policy & Practice</i> . Vol.15, No.3, pp.245-251. |
| 1251 | M | Crossley, M (2008) Boundaries and diversity in comparative education. <i>Comparative Education</i> . Vol.44, No.4, pp.371- |
| 1252B | B | Takayama, K (2008) The politics of international league tables: PISA in Japan's achievement crisis debate. <i>Comparative Education</i> . Vol.44, No.4, pp.387-407. |
| 1253A | A1 | Xu, J (2008) Sibship Size and Educational Achievement: The Role of Welfare Regimes Cross-Nationally. <i>Comparative Education Review</i> . Vol.52, No.3, pp.413-436. |
| 1254A | A1 | Ma, X (2008) Within-School Gender Gaps in Reading, Mathematics, and Science Literacy. <i>Comparative Education Review</i> . Vol.52, No.3, pp.437-460. |
| 1255A | A1 | Dronkers, J & Robert, P (2008) Differences in Scholastic Achievement of Public, Private Government-Dependent, and Private Independent Schools: A Cross-National Analysis. <i>Educational Policy</i> . Vol.22, No.4, pp.541-577. |
| 1256A | A2 | Hatzinikita, V, Dimopoulos, K & Christidou, V (2008) PISA test items and school textbooks related to science: A textual comparison. <i>Science Education</i> . Vol.92, No.4, pp.664-687. |
| 1257B | B | Saha, L J & Sikora, J (2008) The Career Aspirations and Expectations of School Students: From Individual to Global Effects. <i>Education & Society</i> . Vol.26, No.2, pp.5-22. |
| 1258 | M | Pfeifer, M & Holtappels, H G (2008) Improving Learning in All-Day Schools: results of a new teaching time model. <i>European Educational Research Journal</i> . Vol.7, No.2, pp.6- |
| 1259 | M | Spring, J (2008) Research on Globalization and Education. <i>Review of Educational Research</i> . Vol.79, No.2, pp.330-363. |
| 1260A | A1 | Hyunjoon, P (2008) The Varied Educational Effects of Parent-Child Communication: A Comparative Study of Fourteen Countries. <i>Comparative Education Review</i> . Vol.52, No.2, pp.219-244. |
| 1261A | A1 | Dupriez, V, Dumay, X & Vause, A (2008) How Do School Systems Manage Pupils' Heterogeneity? <i>Comparative Education Review</i> . Vol.52, No.2, pp.245-273. |
| 1262 | D | |
| 1263B | B | Simola, H (2008) Understanding the Miracle of PISA: Historical and Sociological Remarks on Finnish Teacherhood. <i>Education & Society</i> . Vol.26, No.1, pp.5-23. |
| 1264 | M | Metz, S (2008) Science for All. <i>Science Teacher</i> . Vol.75, No.3, pp.8- |
| 1265A | A1 | Jenkins, S P, Micklewright, J & Schnepf, S V (2008) Social segregation in secondary schools: how does England compare with other countries? <i>Oxford Review of Education</i> . Vol.34, No.1, pp.21-37. |
| 1266A | A1 | Marks, G N (2008) Accounting for the gender gaps in student performance in reading and mathematics: evidence from 31 countries. <i>Oxford Review of Education</i> . Vol.34, No.1, pp.89-109. |
| 1267A | A2 | Bassani, C (2008) Social Capital and Disparities in Canadian Youth's Mathematics Achievement. <i>Canadian Journal of Education/Revue canadienne de l'éducation</i> . Vol.31, No.3, pp.727-760. |
| 1268A | A2 | Edgerton, J D, Peter, T & Roberts, L W (2008) Back to the Basics: Socio-Economic, Gender, and Regional Disparities in Canada's Educational System. <i>Canadian Journal of Education/Revue canadienne de l'éducation</i> . Vol.31, No.4, pp.861-888. |
| 1269C | C | Nonoyama-Tarumi, Y (2008) Cross-National Estimates of the Effects of Family Background on Student Achievement: A Sensitivity Analysis. <i>International Review of Education/Internationale Zeitschrift für Erziehungswissenschaft/Revue Internationale de l'Education</i> . Vol.54, No.1, pp.57-82. |
| 1270B | B | Rautalin, M & Alasuutari, P (2007) The Curse of Success: the impact of the OECD's Programme for International Student Assessment on the discourses of the teaching profession in Finland. <i>European</i> |

| # | C | Reference |
|-------|----|--|
| | | <i>Educational Research Journal</i> . Vol.6, No.4, pp.3- |
| 1271A | A1 | Brozo, W G, Shiel, G & Topping, K (2008) Engagement in reading: Lessons learned from three PISA countries. <i>Journal of Adolescent & Adult Literacy</i> . Vol.51, No.4, pp.304-315. |
| 1272 | N | Science rocks. (2007) <i>OECD Observer</i> . No.264/265, p.21. |
| 1273A | A2 | Marks, G N (2007) Do Schools Matter for Early School Leaving? Individual and School Influences in Australia. <i>School Effectiveness & School Improvement</i> . Vol.18, No.4, pp.429-450. |
| 1274B | B | Takayama, K (2007) "A Nation at Risk" Crosses the Pacific: Transnational Borrowing of the U.S. Crisis Discourse in the Debate on Education Reform in Japan. <i>Comparative Education Review</i> . Vol.51, No.4, pp.423-446. |
| 1275 | N | Document Digest. (2007) <i>Education Journal</i> . No.106, p.18. |
| 1276A | A2 | Samuelstuen, M S, Bråten, I & Valås, H (2007) Context Effects in Norwegian 10th-Grade Students' Reports on Learning Strategies using the Cross-Curricular Competencies Instrument. <i>Scandinavian Journal of Educational Research</i> . Vol.51, No.5, pp.511-529. |
| 1277 | M | Baybee, R W (2007) Science teaching and international assessments. <i>Science Teacher</i> . Vol.74, No.8, pp.41-48. |
| 1278 | D | |
| 1279A | NA | Maslowski, R, Scheerens, J & Luyten, H (2007) The effect of school autonomy and school internal decentralization on students' reading literacy. <i>School Effectiveness & School Improvement</i> . Vol.18, No.3, pp.303-334. |
| 1280 | M | Van Petegem, P & Vanhoof, J (2007) Towards a model of effective school feedback: School heads' points of view. <i>Educational Research & Evaluation</i> . Vol.13, No.4, pp.311-325. |
| 1281A | A2 | Ammermueller, A (2007) Poor Background or Low Returns? Why Immigrant Students in Germany Perform so Poorly in the Programme for International Student Assessment. <i>Education Economics</i> . Vol.15, No.2, pp.215-230. |
| 1282 | M | Below, S (2007) What are the chances of young Turks and Italians for equal education and employment in Germany? The role of objective and subjective indicators. <i>Social Indicators Research</i> . Vol.82, No.2, pp.209-231. |
| 1283 | D | |
| 1284A | A2 | Kivirauma, J & Ruoho, K (2007) Excellence through Special Education? Lessons from the Finnish School Reform. <i>International Review of Education/Internationale Zeitschrift für Erziehungswissenschaft</i> . Vol.53, No.3, pp.283-302. |
| 1285 | M | Hill, I (2007) Multicultural and International Education: Never the Twain Shall Meet? <i>International Review of Education/Internationale Zeitschrift für Erziehungswissenschaft/Revue Internationale de l'Education</i> . Vol.53, No.3, pp.245-264. |
| 1286 | M | Treagust, D F & Chandrasegaran, A L (2007) The Taiwan National Science Concept Learning Study in an International Perspective. <i>International Journal of Science Education</i> . Vol.29, No.4, pp.391-403. |
| 1287 | M | Prais, S J (2007) Two recent (2003) international surveys of schooling attainments in mathematics: England's problems. <i>Oxford Review of Education</i> . Vol.33, No.1, pp.33-46. |
| 1288A | A2 | Ho, E S-C (2007) Association between Self-Related Cognition and Mathematics Performance: The Case in Hong Kong. <i>Education Journal</i> . Vol.35, No.2, pp.59-76. |
| 1289 | M | Sammons, P (2006) The contribution of international studies on Educational Effectiveness: Current and future directions. <i>Educational Research & Evaluation</i> . Vol.12, No.6, pp.583-593. |
| 1290 | M | Marsh, H W, Kit-Tai, H, Artelt, C, Baumert, J & Peschar, J L (2006) OECD's Brief Self-Report Measure of Educational Psychology's Most Useful Affective Constructs: Cross-Cultural, Psychometric Comparisons Across 25 Countries. <i>International Journal of Testing</i> . Vol.6, No.4, pp.311-360. |
| 1291 | M | Phillips, D (2006) Investigating policy attraction in education. <i>Oxford Review of Education</i> . Vol.32, No.5, pp.551-559. |
| 1292B | B | Ertl, H (2006) Educational standards and the changing discourse on education: the reception and consequences of the PISA study in Germany. <i>Oxford Review of Education</i> . Vol.32, No.5, pp.619-634. |
| 1293A | A2 | Leino, K (2006) Reading the Web—Students' Perceptions about the Internet. <i>Scandinavian Journal of Educational Research</i> . Vol.50, No.5, pp.541-557. |
| 1294B | B | Huber, S G & Gördel, B (2006) Quality Assurance in the German School System. <i>European Educational Research Journal</i> . Vol.5, No.3/4, p.3- |
| 1295B | B | Hjort, K (2006) De-democratisation in Denmark? <i>European Educational Research Journal</i> . Vol.5, No.3/4, pp.6- |
| 1296A | NA | Van Langen, A, Bosker, R & Dekkers, H (2006) Exploring cross-national differences in gender gaps in education. <i>Educational Research & Evaluation</i> . Vol.12, No.2, pp.155-177. |
| 1297 | NA | Corten, R & Dronkers, J (2006) School achievement of pupils from the lower strata in public, private government-dependent and private government-independent schools: A cross-national test of the Coleman-Hoffer thesis 1. <i>Educational Research & Evaluation</i> . Vol.12, No.2, pp.179-208. |
| 1298B | B | Grubera, K H (2006) The German 'PISA-shock': some aspects of the extraordinary impact of the OECD's PISA study on the German education system. <i>Oxford Studies in Comparative Education</i> . Vol.16, No.1, |

| # | C | Reference |
|-------|----|---|
| | | pp.195-208. |
| 1299A | A1 | Marks, G (2006) Are between- and within-school differences in student performance largely due to socio-economic background? Evidence from 30 countries. <i>Educational Research</i> . Vol.48, No.1, pp.21-40. |
| 1300A | NA | Thorpe, G (2006) Multilevel analysis of PISA 2000 reading results for the United Kingdom using pupil scale variables. <i>School Effectiveness & School Improvement</i> . Vol.17, No.1, pp.33-62. |
| 1301B | B | Stack, M (2006) Testing, Testing, Read All about It: Canadian Press Coverage of the PISA Results. <i>Canadian Journal of Education/Revue canadienne de l'éducation</i> . Vol.29, No.1, pp.49-69. |
| 1302A | A1 | Papanastasiou, E C, Zembylas, M & Vrasidas, C (2005) An Examination of the PISA Database to Explore the Relationship Between Computer Use and Science Achievement. <i>Educational Research & Evaluation</i> . Vol.11, No.6, pp.529-543. |
| 1303A | A1 | Williams, T, Williams, K, Kastberg, D & Jocelyn, L (2005) Achievement and affect in OECD nations. <i>Oxford Review of Education</i> . Vol.31, No.4, pp.517-545. |
| 1304B | B | Simola, H (2005) The Finnish miracle of PISA: historical and sociological remarks on teaching and teacher education. <i>Comparative Education</i> . Vol.41, No.4, pp.455-470. |
| 1305A | A1 | Duru-Bellat, M & Suchaut, B (2005) Organisation and Context, Efficiency and Equity of Educational Systems: what PISA tells us. <i>European Educational Research Journal</i> . Vol.4, No.3, pp.181-194. |
| 1306 | N | Achievement in Science. (2005) <i>Education Journal</i> . No.86, p.20. |
| 1307B | B | Livingston, K & McCall, J (2005) Evaluation: judgemental or developmental? <i>European Journal of Teacher Education</i> . Vol.28, No.2, pp.165-178. |
| 1308B | B | Ferrini-Mundy, J & Schmidt, W H (2005) International Comparative Studies in Mathematics Education: Opportunities for Collaboration and Challenges for Researchers. <i>Journal for Research in Mathematics Education</i> . Vol.36, No.3, pp.164-175. |
| 1309 | D | |
| 1310C | C | Nash, R (2005) Explanation and quantification in educational research: the arguments of critical and scientific realism. <i>British Educational Research Journal</i> . Vol.31, No.2, pp.185-204. |
| 1311 | M | Sfard, A (2005) What Could be More Practical than Good Research? <i>Educational Studies in Mathematics</i> . Vol.58, No.3, pp.393-413. |
| 1312 | M | Geissler, G (2005) The Interrelationship between the State and the Local Community in the German School Tradition. <i>European Education</i> . Vol.37, No.1, pp.62-82. |
| 1313B | B | Baker, R & Jones, A (2005) How can international studies such as the international mathematics and science study and the programme for international student assessment be used to inform practice, policy and future research in science education in New Zealand? <i>International Journal of Science Education</i> . Vol.27, No.2, pp.145-157. |
| 1314A | A2 | Matrebreveju, P & Strakova, J (2005) The role of the family and the school in the reproduction of educational inequalities in the post-Communist Czech Republic. <i>British Journal of Sociology of Education</i> . Vol.26, No.1, pp.17-40. |
| 1315 | D | |
| 1316 | D | |
| 1317A | NA | Pratt, H (2005) Where Are We Now? <i>Science Teacher</i> . Vol.72, No.1, pp.10-11. |
| 1318 | N | PISA. (2004) <i>Education Journal</i> . No.81, p.8. |
| 1319 | N | Hanushek, E A (2004) Our School Performance Matters. <i>Journal of Education</i> . Vol.185, No.3, pp.1-6. |
| 1320C | C | Goldstein, H (2004) International comparisons of student attainment: some issues arising from the PISA study. <i>Assessment in Education: Principles, Policy & Practice</i> . Vol.11, No.3, pp.319-330. |
| 1321 | M | Georg, W (2004) Cultural Capital and Social Inequality in the Life Course. <i>European Sociological Review</i> . Vol.20, No.4, pp.333-344. |
| 1322B | B | Lie, S & Linnakylä, P (2004) Nordic PISA 2000 in a sociocultural perspective. <i>Scandinavian Journal of Educational Research</i> . Vol.48, No.3, pp.227- |
| 1323A | NA | Taube, K (2004) Factors behind low reading literacy achievement. <i>Journal of Educational Research</i> . Vol.48, No.3, pp.231-249. |
| 1324A | A2 | Leino, K, Linnakylä, P & Malin, A (2004) Finnish students' multiliteracy profiles. <i>Scandinavian Journal of Educational Research</i> . Vol.48, No.3, pp.251-270. |
| 1325A | A1 | Kjaernsli, M & Lie, S (2004) PISA and scientific literacy: similarities and differences between the Nordic countries. <i>Scandinavian Journal of Educational Research</i> . Vol.48, No.3, pp.271-286. |
| 1326A | A1 | Turmo, A (2004) Scientific literacy and socio-economic background among 15-year-olds-a Nordic perspective. <i>Scandinavian Journal of Educational Research</i> . Vol.48, No.3, pp.287-305. |
| 1327A | A2 | Hvistendahl, R & Roe, A (2004) The literacy achievement of Norwegian minority students. <i>Scandinavian Journal of Educational Research</i> . Vol.48, No.3, pp.307-324. |
| 1328C | C | Olsen, R V (2004) The search for descriptions of students' thinking and knowledge: exploring nominal cognitive variables by correspondence and homogeneity analysis. <i>Scandinavian Journal of Educational Research</i> . Vol.48, No.3, pp.325-341. |
| 1329A | A1 | Kinnakylä, P, Malin, A & Taube, K (2004) Factors behind Low Reading Literacy Achievement. <i>Scandinavian Journal of Educational Research</i> . Vol.48, No.3, pp.231-249. |

| # | C | Reference |
|-------|----|--|
| 1330A | NA | Marsh, H W & Kit-Tai, H (2004) Explaining Paradoxical Relations Between Academic Self-Concepts and Achievements: Cross-Cultural Generalizability of the Internal/External Frame of Reference Predictions Across 26 Countries. <i>Journal of Educational Psychology</i> . Vol.96, No.1, pp.56-67. |
| 1331 | N | McGaw, B (2004) Quality education. <i>OECD Observer</i> . No.242, p.9-10. |
| 1332 | N | McKenzie, P (2004) Teaching. <i>OECD Observer</i> . No.242, pp.11-13. |
| 1333 | D | |
| 1334A | NA | Ho, E S-C (2004) Self-Regulated Learning and Academic Achievement of Hong Kong Secondary School Students. <i>Education Journal</i> . Vol.32, No.2, pp.87-107. |
| 1335 | N | Lingens, H (2003) Editor's Introduction. <i>European Education</i> . Vol.35, No.4, pp.3-6. |
| 1336 | M | Schröder, G R (2003) One Law for All Schools. <i>European Education</i> . Vol.35, No.4, pp.13-20. |
| 1337B | B | Bulmahn, E (2003) National Education Standards as Part of a Comprehensive Quality Management System for Schools. <i>European Education</i> . Vol.35, No.4, pp.21-25. |
| 1338 | M | Klemm, K (2003) If No Reform Then at Least No Rollback. <i>European Education</i> . Vol.35, No.4, pp.34-43. |
| 1339B | B | Schmidt, G (2003) Reactions of Participating Countries as Reflected in Their Press. <i>European Education</i> . Vol.35, No.4, pp.58-69. |
| 1340 | M | Von Kopp, B (2003) On the Question of Cultural Context as a Factor in International Academic Achievement. <i>European Education</i> . Vol.35, No.4, pp.70-98. |
| 1341A | A2 | Nash, R (2003) Is the School Composition Effect Real?: A Discussion With Evidence From the UK PISA Data. <i>School Effectiveness & School Improvement</i> . Vol.14, No.4, pp.441-457. |
| 1342A | A1 | Papanastasiou, E C (2003) Science Literacy by Technology by Country: USA, Finland and Mexico. Making sense of it all. <i>Research in Science & Technological Education</i> . Vol.21, No.2, pp.229-241. |
| 1343A | A2 | Papanastasiou, E C, Zembylas, M & Vrasidas, C (2003) Can Computer Use Hurt Science Achievement? The USA Results from PISA. <i>Journal of Science Education and Technology</i> . Vol.12, No.3, pp.325-332. |
| 1344C | C | Adams, R J (2003) Response to 'Cautions on OECD's Recent Educational Survey (PISA)'. <i>Oxford Review of Education</i> . Vol.29, No.3, pp.377- |
| 1345 | N | Brazil, Chile, Peru students struggle. (2003) <i>OECD Observer</i> . No.238, p.2/5. |
| 1346 | N | Diploma of excellence. (2003) <i>OECD Observer</i> . No.237, p.6. |
| 1347 | N | PISA. (2003) <i>Education Journal</i> . No.66, p.15. |
| 1348A | A2 | Ho, E S-C (2003) Accomplishment and Challenges of Hong Kong Education System: What We Have Learned from PISA. <i>Education Journal</i> . Vol.32, No.2, pp.91-120. |
| 1349A | A2 | Wong, K-M P (2003) Mathematical Literacy of Hong Kong's 15-Year-Old Students in PISA. <i>Education Journal</i> . Vol.31, No.2, pp.91-120. |
| 1350A | A2 | Chun, C K-W, Tong, C-W & Sze, P M-M (2003) Reading Performance of Hong Kong's 15-Year-Old Students in PISA 2000. <i>Education Journal</i> . Vol.31, No.2, pp.57-89. |
| 1351A | NA | Wu, M (2003) The Impact of PISA on Mathematics Education: Linking Mathematics and the Real World. <i>Education Journal</i> . Vol.31, No.2, pp.121-140. |
| 1352A | NA | Valijärvi, J (2003) The System and How Does It Work: Some Curricular and Pedagogical Characteristics of the Finnish Comprehensive School. <i>Education Journal</i> . Vol.31, No.2, pp.31-55. |
| 1353A | NA | Willms, J D (2003) What Can We Say about the Quality and Equality of Educational Systems from the First Cycle of the PISA? <i>Education Journal</i> . Vol.31, No.2, pp.161-175. |
| 1354 | M | Bonnet, G (2002) Reflections in a Critical Eye: on the pitfalls of international assessment. <i>Assessment in Education: Principles, Policy & Practice</i> . Vol.9, No.3, pp.387-399. |
| 1355 | N | Mulford, B (2002) Sorting the Wheat from the Chaff – Knowledge and Skills for Life: first results from OECD's PISA 2000. <i>European Journal of Education</i> . Vol.37, No.2, p.211. |
| 1356 | N | Bulmahn, E (2002) PISA The consequences for Germany. <i>OECD Observer</i> . No.231/232, p.33. |
| 1357 | N | Hirsch, D (2002) How good is our global education? The PISA survey. <i>OECD Observer</i> . No.230, p.37. |
| 2001A | A2 | Liou, P-Y (2014) Examining the Big-Fish-Little-Pond Effect on Students' Self-Concept of Learning Science in Taiwan Based on the TIMSS Databases. <i>International Journal of Science Education</i> . Vol.36, No.12, pp.2009-20028. |
| 2002A | A1 | Nixon, R S & Barth, K N (2014) A Comparison of TIMSS Items Using Cognitive Domains. <i>School Science & Mathematics</i> . Vol.114, No.2, pp.65-75. |
| 2003A | A2 | De Witte, K & Van Klaveren, C (2014) How are teachers teaching? A nonparametric approach. <i>Education Economics</i> . Vol.22, No.1, pp.3-23. |
| 2004A | A2 | Breton, T R (2014) Evidence that class size matters in 4th grade mathematics: An analysis of TIMSS 2007 data for Colombia. <i>International Journal of Educational Development</i> . Vol.34, pp.51-57. |
| 2005A | A1 | Jerrim, J (2014) The Unrealistic Educational Expectations of High School Pupils: Is America Exceptional? <i>Sociological Quarterly</i> . Vol.55, No.1, pp.196-231. |
| 2006A | A2 | Mathematics instruction and achievement of eight-grade students in Korea: Results from the TIMSS 2007 Assessment (2013) <i>Education</i> . Vol.134, No.2, pp.266-270. |
| 2007C | C | DeMars, C E (2013) A Tutorial on Interpreting Bifactor Model Scores. <i>International Journal of Testing</i> . Vol.12, No.4, pp.354-378. |
| 2008B | B | Kind, P M (2013) Conceptualizing the Science Curriculum: 40 Years of Developing Assessment |

| # | C | Reference |
|-------|----|--|
| | | Frameworks in Three Large-Scale Assessments. <i>Science Education</i> . Vol.97, No.5, pp.671-694. |
| 2009A | A2 | De Simone, G (2013) Render unto primary the things which are primary's: Inherited and fresh learning divides in Italian lower secondary education. <i>Economics of Education Review</i> . Vol.35, pp.12-23. |
| 2010A | NA | Rutkowski, L, Rutkowski, D & Engel, L (2013) Sharp Contrasts at the Boundaries: School Violence and Educational Outcomes Internationally. <i>Comparative Education Review</i> . Vol.57, No.2, pp.232-259. |
| 2011A | A2 | Mills, J D & Holloway, C E (2013) The development of statistical literacy skills in the eighth grade: exploring the TIMSS data to evaluate student achievement and teacher characteristics in the United States. <i>Educational Research & Evaluation</i> . Vol.19, No.4, pp.323-245. |
| 2012A | NA | Pahlke, E, Hyde, J S & Mertz, J E (2013) The Effects of Single-Sex Compared With Coeducational Schooling on Mathematics and Science Achievement: Data From Korea. <i>Journal of Educational Psychology</i> . Vol.105, No.2, pp.444-452. |
| 2013C | NA | Larson, M R & Leinwand, S (2013) Prepare for More Realistic Testing Results. <i>Mathematics Teaching in the Middle School</i> . Vol.18, No.9, pp.524-526. |
| 2014 | D | |
| 2015A | NA | Marsh, H W, Abu-Hilal, M M, Abdelfattah, F, Xu, M K, Abduljabbar, A S, Morin, A J S, Leung, K C, Nagengast, B & Parker, P (2013) Factorial, Convergent, and Discriminant Validity of TIMSS Math and Science Motivation Measures: A Comparison of Arab and Anglo-Saxon Countries. <i>Journal of Educational Psychology</i> . Vol.105, No.1, pp.108-128. |
| 2016A | A2 | Shymansky, J, Annetta, L, Yore, L D, Wang, T-L & Everett, S (2013) The Impact of a Multiyear Systemic Reform Effort on Rural Elementary School Students' Science Achievement. <i>School Science & Mathematics</i> . Vol.113, No.2, pp.69-79. |
| 2017 | D | |
| 2018 | N | East Asian Students Still Excel in Math, Science. (2012) <i>Science</i> . Vol.338, No.6114, p.1516. |
| 2019A | A2 | Sezer, R, Güner, N & Akkus, I O (2012) Teachers' perspective on whether the mathematics reform will change Turkey's ranking in TIMSS. <i>Education</i> . Vol.133, No.2, pp.391-411. |
| 2020A | A1 | Yu, C H (2012) Examining the relationships among academic self-concept, instrumental motivation, and TIMSS 2007 science scores: a cross-cultural comparison of five East Asian countries/regions and the United States. <i>Educational Research & Evaluation</i> . Vol.18, No.8, pp.713-731. |
| 2021A | A2 | Schmidt, W H & Houang, R T (2012) Curricular Coherence and the Common Core State Standards for Mathematics. <i>Educational Researcher</i> . Vol.41, No.8, pp.294-308. |
| 2022A | A1 | Mesic, V (2012) Identifying Country-Specific Cultures of Physics Education: A differential item functioning approach. <i>International Journal of Science Education</i> . Vol.34, No.16, pp.2483-2500. |
| 2023A | A2 | Mohammadpour, E (2012) A multilevel study on trends in Malaysian secondary school students' science achievement and associated school and student predictors. <i>Science Education</i> . Vol.96, No.6, pp.1013-1046. |
| 2024A | A2 | Chen, S-F, Lin, C-Y, Wang, J-R, Lin, S-W & Kao, H-L (2012) A Cross-grade Comparison to Examine the Context Effect on the Relationships Among Family Resources, School Climate, Learning Participation, Science Attitude, and Science Achievement Based on TIMSS 2003 in Taiwan. <i>International Journal of Science Education</i> . Vol.34, No.14, pp.2089-2106. |
| 2025A | A1 | Helbig, M (2012) Boys do not benefit from male teachers in their reading and mathematics skills: empirical evidence from 21 European Union and OECD countries. <i>British Journal of Sociology of Education</i> . Vol.33, No.5, pp.661-677. |
| 2026A | A2 | Hansson, Å (2012) The meaning of mathematics instruction in multilingual classrooms: analyzing the importance of responsibility for learning. <i>Educational Studies in Mathematics</i> . Vol.81, No.1, pp.103-125. |
| 2027A | NA | House, J D (2012) Science Achievement of Elementary-School Students in the United States and Japan in TIMSS 2007: An Assessment of the Effects of Technology Engagement and Classroom Lesson Activities. <i>International Journal of Instructional Media</i> . Vol.39, No.3, pp.263-274. |
| 2028A | NA | Vandecandelaere, M, Speybroeck, S, Vanlaar, G, De Fraine, B & Van Damme, J (2012) Learning environment and students' mathematics attitude. <i>Studies in Educational Evaluation</i> . Vol.38, No.3/4, pp.107-120. |
| 2029A | A1 | Wiseman, A W (2012) The Impact of Student Poverty on Science Teaching and Learning: A Cross-National Comparison of the South African Case. <i>American Behavioral Scientist</i> . Vol.56, No.7, pp.941-960. |
| 2030A | A1 | Cho, I (2012) The effect of teacher-student gender matching: Evidence from OECD countries. <i>Economics of Education Review</i> . Vol.31, No.3, pp.54-67. |
| 2031A | NA | House, J D (2012) Motivational Effects of Computers and Classroom Instruction for Science Learning: An Examination of eight-grade students in the United States and Korea in the TIMSS 2007 Assessment. <i>International Journal of Instructional Media</i> . Vol.39, No.2, pp.169-179. |
| 2132A | A2 | Chen, Y-H (2012) Cognitive diagnosis of mathematics performance between rural and urban students in Taiwan. <i>Assessment in Education: Principles, Policy & Practice</i> . Vol.19, No.2, pp.193-209. |
| 2133A | NA | House, J D & Telese, J A (2012) Effects of Mathematics Lesson Activities and Computer Use on Algebra Achievement of Eighth-Grade Students in the United States and Japan: findings from the TIMSS 2007 assessment. <i>International Journal of Instructional Media</i> . Vol.39, No.1, pp.69-81. |
| 2134A | A1 | Dodeen, H, Abdelfattah, F, Shumrani, S & Hilal M A (2012) The Effects of Teachers' Qualifications, |

| # | C | Reference |
|-------|----|---|
| | | Practices, and Perceptions on Student Achievement in TIMSS Mathematics: A Comparison of Two Countries. <i>International Journal of Testing</i> . Vol.12, No.1, pp.61-77. |
| 2035 | M | Bulle, N (2011) Comparing OECD educational models through the prism of PISA. <i>Comparative Education</i> . Vol.47, No.4, pp.503-521. |
| 2036C | C | Wang, J (2011) Re-examining Test Item Issues in the TIMSS Mathematics and Science Assessments. <i>School Science & Mathematics</i> . Vol.111, No.7, pp.334-344. |
| 2037C | C | Eisenkopf, G (2011) Paying for better test scores. <i>Education Economics</i> . Vol.19, No.4, pp.329-339. |
| 2038A | NA | House, J D & Telese, J A (2011). Effects of Computer Activities and Classroom Lesson Strategies on Motivation for Mathematics Learning for Eighth-Grade Students in the United States and Korea. <i>International Journal of Instructional Media</i> . Vol.38, No.3, pp.295-306. |
| 2039A | A2 | Santau, A O, Maerten-Rivera, J L & Huggins, A C (n.d.) Science Achievement of English Language Learners in Urban Elementary Schools: Fourth-Grade Student Achievement Results from a Professional Development Intervention. <i>Science Education</i> . Vol.95, No.5, pp.771-793. |
| 2040A | NA | House, J D (2011) Effects of Computer Activities and Classroom Instructional Strategies on Science Achievement of Eighth-Grade Students in the United States and Korea: Results from the TIMSS 2007 Assessment. <i>International Journal of Instructional Media</i> . Vol.38, No.2, pp.197-208. |
| 2041B | NA | Schmidt, W H, Cogan, L S, Houang, R T & McKnight, CC (2011) Content Coverage Differences across Districts/States: A Persisting Challenge for U.S. Education Policy. <i>American Journal of Education</i> . Vol.117, No.3, pp.399-427. |
| 2042A | A2 | Young-Sun, L, Yoon, S P & Taylan, D (2011) A Cognitive Diagnostic Modeling of Attribute Mastery in Massachusetts, Minnesota, and the U.S. National Sample Using the TIMSS 2007. <i>International Journal of Testing</i> . Vol.11, No.2, pp.144-177. |
| 2043A | NA | Jane-Jane, L & Feng-Chiu, T (2011) Taiwanese Arithmetic and Algebra. <i>Mathematics Teaching in the Middle School</i> . Vol.16, No.7, pp.422-429. |
| 2044 | D | |
| 2045 | M | Bybee, R & McCrae, B (2011) Scientific Literacy and Student Attitudes: Perspectives from PISA 2006 science. <i>International Journal of Science Education</i> . Vol.33, No.1, pp.7-26. |
| 2046 | D | |
| 2047A | A2 | Hansson, Å (2010) Instructional responsibility in mathematics education: modelling classroom teaching using Swedish data. <i>Educational Studies in Mathematics</i> . Vol.75, No.2, pp.171-189. |
| 2048A | A1 | Hueveline, P, Hongxing, Y & Timberlake, J M (2010) It Takes a Village (Perhaps a Nation): Families, States, and Educational Achievement. <i>Journal of Marriage & Family</i> . Vol.72, No.5, pp.1362-1376. |
| 2049A | A2 | Yi-Hsin, C, Ferron, J M, Thompson, M S, Gorin, J S, Tatsuoka, K K (2010) Group comparisons of mathematics performance from a cognitive diagnostic perspective. <i>Educational Research & Evaluation</i> . Vol.16, No.4, pp.325-343. |
| 2050A | NA | Kaya, S & Rice, D C (2010) Multilevel Effects of Student and Classroom Factors on Elementary Science Achievement in Five Countries. <i>International Journal of Science Education</i> . Vol.32, No.10, pp.1337-1363. |
| 2051B | NA | Van de Werfhorst, H G & Mijs, J J B (2010) Achievement Inequality and the Institutional Structure of Educational Systems: A Comparative Perspective. <i>Annual Review of Sociology</i> . Vol.36, pp.407-428. |
| 2052A | A1 | Im, S & Park, H J (2010) A comparison of US and Korean students' mathematics skills using a cognitive diagnostic testing method: linkage to instruction. <i>Educational Research & Evaluation</i> . Vol.16, No.3, pp.287-301. |
| 2053A | A2 | Elsaleh, I (2010) Research in Brief: Teachers' Interactions with Curriculum Materials in Mathematics. <i>School Science & Mathematics</i> . Vol.110, No.4, pp.177-179. |
| 2054A | NA | House, J D (2010) Effects of Computer Activities and Instructional Strategies on Science Achievement of Students in the United States and Japan: Results from the TIMSS 2003 Assessment. <i>International Journal of Instructional Media</i> . Vol.37, No.1, pp.103-114. |
| 2055B | B | Chen, S-S & Luoh, M-C (2010) Are Mathematics and Science Test Scores Good Indicators of Labor-Force Quality? <i>Social Indicators Research</i> . Vol.96, No.1, pp.133-143. |
| 2056A | NA | 'Cross-national patterns of gender differences in mathematics: A meta-analysis' Correction. (2010) <i>Psychological Bulletin</i> . Vol.136, No.2, pp.301. |
| 2057A | NA | Capraro, R M, Capraro, M M, Yetkiner, Z E, Ozel, S, Kim H G & Küçük, A R (2010) An international comparison of grade 6 students' understanding of the equal sign. <i>Psychological Reports</i> . Vol.106, No.1, pp.49-53. |
| 2058B | B | Leighton, J P, Gokiert, R J, Cor, M K & Heffernan, C (2010) Teacher beliefs about the cognitive diagnostic information of classroom- versus large-scale tests: implications for assessment literacy. <i>Assessment in Education: Principles, Policy & Practice</i> . Vol.17, No.1, pp.7-21. |
| 2059C | C | Cliffordson, C (2010) Methodological issues in investigations of the relative effects of schooling and age on school performance: the between-grade regression discontinuity design applied to Swedish TIMSS 1995 data. <i>Educational Research & Evaluation</i> . Vol.16, No.1, pp.39-52. |
| 2060B | B | Barry, C L, Horst, S J & Finney, S J (n.d.) Do Examinees Have Similar Test-Taking Effort? A High-Stakes Question for Low-Stakes Testing. <i>International Journal of Testing</i> . Vol.10, No.4, pp.342-363. |

| # | C | Reference |
|-------|----|---|
| 2061A | A1 | Else-Quest, N M, Hyde, J S & Linn, M C (2010) Cross-National Patterns of Gender Differences in Mathematics: A Meta-Analysis. <i>Psychological Bulletin</i> . Vol.136, No.1, pp.103-127. |
| 2062 | D | |
| 2063A | A2 | House, J D (2009) Elementary-school Mathematics Instruction and Achievement of fourth-grade Students in Japan: Findings from the TIMSS 2007 Assessment. <i>Education</i> . Vol.130, No.2, pp.301-307. |
| 2064A | NA | Andrews, P (2009) Mathematics Teachers' Didactic Strategies: Examining the Comparative Potential of Low Inference Generic Descriptors. <i>Comparative Education Review</i> . Vol.53, No.4, pp.559-581. |
| 2065A | NA | Binder, M (2009) Why Are Some Low-Income Countries Better at Providing Secondary Education? <i>Comparative Education Review</i> . Vol.53, No.4, pp.513-534. |
| 2066A | NA | Espinoza, F (2009) Using Project-Based Data in Physics to Examine Television Viewing in Relation to Student Performance in Science. <i>Journal of Science Education and Technology</i> . Vol.18, No.5, p.458-465. |
| 2067A | A1 | Chudgar, A & Luschei, T F (2009) National Income, Income Inequality, and the Importance of Schools: A Hierarchical Cross-National Comparison. <i>American Educational Research Journal</i> . Vol.46, No.3, pp.626-658. |
| 2068A | NA | Own, Z-Y, Ling, F-N & Chuang, K-H (2009) The Study and Application of the web based Constructed Learning Theory of a "Chemical reaction". <i>International Journal of Instructional Media</i> . Vol.36, No.3, pp.339-350. |
| 2069A | NA | Min-Hsiung, H (2009) Beyond horse race comparisons of national performance averages: math performance variation within and between classrooms in 38 countries. <i>Educational Research & Evaluation</i> . Vol.15, No.4, pp.327-342. |
| 2070 | N | Jasny, B R (2009) Correlates of the Gender Gap. <i>Science</i> . Vol.325, No.5938, pp.249. |
| 2071A | A2 | Schechtman, E & Yitzhaki, S (2009) Ranking Groups' Abilities: Is it Always Reliable? <i>International Journal of Testing</i> . Vol.9, No.3, pp.195-214. |
| 2072A | NA | House, J D (2009) Mathematics beliefs and achievement of a national sample of Native American students: results from the Trends in International Mathematics and Science Study (TIMSS) 2003 United States assessment. <i>Psychological Reports</i> . Vol.104, No.2, pp.439-446. |
| 2073B | B | Rutkowski, L & Rutkowski, D (2009) Trends in TIMSS responses over time: evidence of global forces in education? <i>Educational Research & Evaluation</i> . Vol.15, No.2, pp.137-152. |
| 2074C | C | Solano-Flores, G, Backhoff, E & Contreras-Nino, L A (2009) Theory of Test Translation Error. <i>International Journal of Testing</i> . Vol.9, No.2, pp.78-91. |
| 2075A | A2 | House, J D (2009) Classroom Instructional Strategies and Science Career Interest for Adolescent Students in Korea: Results from the TIMSS 2003 Assessment. <i>Journal of Instructional Psychology</i> . Vol.36, No.1, pp.13-19. |
| 2076A | A2 | House, J D (2008) Effects of Classroom Instructional Strategies and Self-Beliefs on Science Achievement of Elementary-School Students in Japan: Results from the TIMSS 2003 Assessment. <i>Education</i> . Vol.129, No.2, pp.259-266. |
| 2077 | M | Huntley, M A & Davis, J D (2008) High-School Students' Approaches to Solving Algebra Problems that are Posed Symbolically: Results from an Interview Study. <i>School Science & Mathematics</i> . Vol.108, No.8, pp.380-388. |
| 2078A | A1 | Penner, A M (2008) Gender Differences in Extreme Mathematical Achievement: An International Perspective on Biological and Social Factors. <i>American Journal of Sociology</i> . Vol.114, pp.138-170. |
| 2079A | A2 | Dogan, E & Tatsuoka, K (2008) An International Comparison Using a Diagnostic Testing Model: Turkish Students' Profile of Mathematical Skills on TIMSS-R. <i>Educational Studies in Mathematics</i> . Vol.68, No.3, pp.263-272. |
| 2080A | NA | Murdock, J (2008) Comparison of Curricular Breadth, Depth, and Recurrence and Physics Achievement of TIMSS Population 3 Countries. <i>International Journal of Science Education</i> . Vol.30, No.9, pp.1135-1157. |
| 2081C | C | Yi-Hsin, C, Gorin, J S, Thompson, M S & Tatsuoka, K K (2008) Cross-Cultural Validity of the TIMSS-1999 Mathematics Test: Verification of a Cognitive Model. <i>International Journal of Testing</i> . Vol.8, No.3, pp.251-271. |
| 2082 | D | |
| 2083A | A1 | Hauger, J B & Sireci, S G (2008) Detecting Differential Item Functioning across Examinees Tested in Their Dominant Language and Examinees Tested in a Second Language. <i>International Journal of Testing</i> . Vol.8, No.3, pp.237-250. |
| 2084A | A1 | Mei-Shiu, C (2008) Achievements and self-concepts in a comparison of math and science: exploring the internal/external frame of reference model across 28 countries. <i>Educational Research & Evaluation</i> . Vol.14, No.3, pp.235-254. |
| 2085B | B | Spring, J (2008) Research on Globalization and Education. <i>Review of Educational Research</i> . Vol.78, No.2, pp.330-363. |
| 2086B | A1 | Neuschmidt, O, Barth, J & Hastedt, D (2008) Trends in gender differences in mathematics and science (TIMSS 1995–2003). <i>Studies in Educational Evaluation</i> . Vol.34, No.2, pp.56-72. |
| 2087A | A2 | Meelissen, M & Luyten, H (2008) The Dutch gender gap in mathematics: Small for achievement, substantial for beliefs and attitudes. <i>Studies in Educational Evaluation</i> . Vol.34, No.2, pp.82-93. |

| # | C | Reference |
|-------|----|--|
| 2088A | A2 | Japelj Pavešič, B (2008) Science achievement, gender differences, and experimental work in classes in Slovenia as evident in TIMSS studies. <i>Studies in Educational Evaluation</i> . Vol.34, No.2, pp.94-104. |
| 2089A | A2 | Zuzovsky, R (2008) Closing achievement gaps between Hebrew-speaking and Arabic-speaking students in Israel: Findings from TIMSS-2003. <i>Studies in Educational Evaluation</i> . Vol.34, No.2, pp.105-117. |
| 2090A | A2 | Howie, S & Scherman, V (2008) The achievement gap between science classrooms and historic inequalities. <i>Studies in Educational Evaluation</i> . Vol.34, No.2, pp.118-130. |
| 2091A | A2 | Thomson, S (2008) Examining the Evidence from TIMSS: Gender Differences in Year 8 Science Achievement in Australia. <i>Studies in Educational Evaluation</i> . Vol.34, No.2, pp.73-81. |
| 2092 | D | |
| 2093C | C | Minkov, M (2008) Self-Enhancement and Self-Stability Predict School Achievement the National Level. <i>Cross-Cultural Research</i> . Vol.42, No.2, pp.172-196. |
| 2094A | A2 | Pugh, G & Telhaj, S (2008) Faith schools, social capital and academic attainment: evidence from TIMSS-R mathematics scores in Flemish secondary schools. <i>British Educational Research Journal</i> . Vol.34, No.2, pp.235-267. |
| 2095A | NA | House, J D (2008) Science Beliefs, Instructional Strategies, and Life Science Achievement of Adolescent Students in Japan: Results from the TIMSS 1999 Assessment. <i>International Journal of Instructional Media</i> . Vol.35, No.1, pp.103-114. |
| 2096B | NA | Hoffman, L R & Brahier, D J (2008) Improving the Planning and Teaching of Mathematics by Reflecting on Research. <i>Mathematics Teaching in the Middle School</i> . Vol.13, No.7, pp.412-417. |
| 2097 | D | |
| 2098A | A2 | Papanastasiou, C (2008) A residual analysis of effective schools and effective teaching in mathematics. <i>Studies in Educational Evaluation</i> . Vol.34, No.1, pp.24-30. |
| 2099 | N | Papanastasiou, C & Plomp, T (2008) Introduction to this special issue. <i>Educational Research & Evaluation</i> . Vol.14, No.1, pp.1-7. |
| 2100A | A1 | Waldfoegel, J & Zhai, F (2008) Effects of public preschool expenditures on the test scores of fourth graders: evidence from TIMSS. <i>Educational Research & Evaluation</i> . Vol.14, No.1, pp.9-28. |
| 2101A | A2 | Howie, S, Scherman, V & Venter, E (2008) The gap between advantaged and disadvantaged students in science achievement in South African secondary schools. <i>Educational Research & Evaluation</i> . Vol.14, No.1, pp.29-46. |
| 2102A | A1 | Kjaernsli, M & Lie, S (2008) Country profiles of scientific competence in TIMSS 2003. <i>Educational Research & Evaluation</i> . Vol.14, No.1, pp.73-85. |
| 2103A | A1 | Shen, C & Tam, H P (2008) The paradoxical relationship between student achievement and self-perception: a cross-national analysis based on three waves of TIMSS data. <i>Educational Research & Evaluation</i> . Vol.14, No.1, pp.87-100. |
| 2104A | A1 | House, J D & Telese, J A (2008) Relationships between student and instructional factors and algebra achievement of students in the United States and Japan: an analysis of TIMSS 2003 data. <i>Educational Research & Evaluation</i> . Vol.14, No.1, pp.101-112. |
| 2105C | C | Andrews, A (2007) Negotiating meaning in cross-national studies of mathematics teaching: kissing frogs to find princes. <i>Comparative Education</i> . Vol.43, No.4, pp.489-509. |
| 2106B | B | Takayama, K (2007) "A Nation at Risk" Crosses the Pacific: Transnational Borrowing of the U.S. Crisis Discourse in the Debate on Education Reform in Japan. <i>Comparative Education Review</i> . Vol.51, No.4, pp.423-446. |
| 2107A | A2 | Dempster, E R & Reddy, V (2007) Item readability and science achievement in TIMSS 2003 in South Africa. <i>Science Education</i> . Vol.91, No.6, pp.906-925. |
| 2108B | NA | Baybee, R W (2007) Science Teaching and International Assessments. <i>Science Teacher</i> . Vol.74, No.8, pp.41-48. |
| 2109 | D | |
| 2110A | A1 | Akiba, M, LeTendre, G K & Scribner, J P (2007) Teacher Quality, Opportunity Gap, and National Achievement in 46 Countries. <i>Educational Researcher</i> . Vol.36, No.7, pp.369-387. |
| 2111 | N | Mervis, J (2007) U.S. Says No to Next Global Test of Advanced Math, Science Students. <i>Science</i> . Vol.317, No.5846, pp.1850. |
| 2112A | NA | House, J D (2007) Cooperative Learning and Computer use during a Geometry Lesson in Japan: A Case Analysis from the TIMSS Videotape Classroom Study. <i>International Journal of Instructional Media</i> . Vol.34, No.3, pp.323-334. |
| 2113A | A2 | Eklöf, H (2007) Test-Taking Motivation and Mathematics Performance in TIMSS 2003. <i>International Journal of Testing</i> . Vol.7, No.3, pp.311-326. |
| 2114A | A2 | Van Petegem, P & Vanhoof, J (2007) Towards a model of effective school feedback: School heads' points of view. <i>Educational Research & Evaluation</i> . Vol.13, No.4, pp.311-325. |
| 2115A | A2 | Hook, W, Bishop, W & Hook, J (2007) A Quality Math Curriculum in Support of Effective Teaching For Elementary Schools. <i>Educational Studies in Mathematics</i> . Vol.65, No.2, pp.125-148. |
| 2116A | NA | House, J D (2007) Relationships Between Self Belief, Instructional Practices, and Chemistry Achievement of Students in Chinese Taipei: Results from the TIMSS 1999 Assessment. <i>International Journal of</i> |

| # | C | Reference |
|-------|----|--|
| | | <i>Instructional Media</i> . Vol.34, No.2, pp.187-206. |
| 2117A | NA | House, J D (2007) Mathematics beliefs and instructional strategies in achievement of elementary-school students in Japan: results from the TIMSS 2003 assessment. <i>Psychological Reports</i> . Vol.100, No.2, pp.476-482. |
| 2118 | M | Treagust, D F & Chandrasegaran, A L (2007) The Taiwan National Science Concept Learning Study in an International Perspective. <i>International Journal of Science Education</i> . Vol.29, No.4, pp.391-403. |
| 2119C | C | Prais, S J (2007) Two recent (2003) international surveys of schooling attainments in mathematics: England's problems. <i>Oxford Review of Education</i> . Vol.33, No.1, pp.33-46. |
| 2120 | D | |
| 2121C | NA | Kyriakides, L (2006) Using international comparative studies to develop the theoretical framework of educational effectiveness research: A secondary analysis of TIMSS 1999 data. <i>Educational Research & Evaluation</i> . Vol.12, No.6, pp.513-534. |
| 2122B | NA | Sammons, P (2006) The contribution of international studies on Educational Effectiveness: Current and future directions. <i>Educational Research & Evaluation</i> . Vol.12, No.6, pp.583-593. |
| 2123A | NA | House, J D (2006) Mathematics Beliefs, Instructional Strategies, and Algebra Achievement of Adolescent Students in Japan: Results from the TIMSS 1999 Assessment. <i>International Journal of Instructional Media</i> . Vol.33, No.4, pp.443-462. |
| 2124A | NA | House, J D (2006) Cognitive-Motivational Characteristics and Geometry Knowledge of Adolescent Students in Japan: Results from the TIMSS 1999 Assessment. <i>International Journal of Instructional Media</i> . Vol.33, No.1, pp.95- |
| 2125 | N | Kulm, G (2006) Reflections on Research in Science and Mathematics Education. <i>School Science & Mathematics</i> . Vol.106, No.8, pp.324-325. |
| 2126A | A2 | Innabi, H & Dodeen, H (2006) Content Analysis of Gender-related Differential Item Functioning TIMSS Items in Mathematics in Jordan. <i>School Science & Mathematics</i> . Vol.106, No.8, pp.328-337. |
| 2127A | A1 | Wu, A D & Ercikan, K (2006) Using Multiple-Variable Matching to Identify Cultural Sources of Differential Item Functioning. <i>International Journal of Testing</i> . Vol.6, No.3, pp.287-300. |
| 2128A | A2 | Xin, M & Papanastasiou, C (2006) How to begin a new topic in mathematics. <i>Evaluation Review</i> . Vol.30, No.4, pp.451-480. |
| 2129C | C | Luyten, H (2006) An empirical assessment of the absolute effect of schooling: regression-discontinuity applied to TIMSS-95. <i>Oxford Review of Education</i> . Vol.32, No.3, pp.397-429. |
| 2130 | D | |
| 2131A | NA | House, J D (2006) The Effects of Classroom Instructional Strategies on Science Achievement of Elementary-School Students in Japan: Findings from the Third International Mathematics and Science Study (TIMSS). <i>International Journal of Instructional Media</i> . Vol.33, No.2, pp.217-229. |
| 2132A | A1 | Barber, N (2006) Is the Effect of National Wealth on Academic Achievement Mediated by Mass Media and Computers? <i>Cross-Cultural Research</i> . Vol.40, No.2, pp.130-151. |
| 2133A | A2 | Liu, X (2006) Student Competence in Understanding the Matter Concept and Its Implications for Science Curriculum Standards. <i>School Science & Mathematics</i> . Vol.106, No.5, pp.220-227. |
| 2134A | A1 | House, J D (2006) Mathematics beliefs and achievement of elementary school students in Japan and the United States: results from the Third International Mathematics and Science Study. <i>The Journal Of Genetic Psychology</i> . Vol.167, No.1, pp.31-45. |
| 2135 | D | |
| 2136A | A2 | Jacobs, J K, Hiebert, J, Givvin, K B, Hollingsworth, H, Garnier, H & Wearne, D (2006) Does Eighth-Grade Mathematics Teaching in the United States Align with the NCTM Standards? Results from the TIMSS 1995 and 1999 Video Studies. <i>Journal for Research in Mathematics Education</i> . Vol.37, No.1, pp.5-32. |
| 2137A | NA | House, J D (2005) Mathematics beliefs and achievement of adolescent students in Japan: results from the TIMSS 1999 assessment. <i>Psychological Reports</i> . Vol.97, No.3, pp.717-720. |
| 2138A | A2 | Törnroos, J (2005) Mathematics textbooks, opportunity to learn and student achievement. <i>Studies in Educational Evaluation</i> . Vol.31, No.4, pp.315-327. |
| 2139B | B | Robertson, I (2005) Issues relating to curriculum, policy and gender raised by national and international surveys of achievement in mathematics. <i>Assessment in Education: Principles, Policy & Practice</i> . Vol.12, No.3, pp.217-236. |
| 2140A | A1 | Desimone, L M, Smith, T, Baker, D & Ueno, K (2005) Assessing Barriers to the Reform of U.S. Mathematics Instruction from an International Perspective. <i>American Educational Research Journal</i> . Vol.42, No.3, pp.501-535. |
| 2141A | A1 | Ammermüller, A, Heijke, H & Wössmann, L (2005) Schooling quality in Eastern Europe: Educational production during transition. <i>Economics of Education Review</i> . Vol.24, No.5, pp.579-599. |
| 2142A | A1 | Leung, F K S (2005) Some Characteristics of East Asian Mathematics Classrooms Based on Data from the Timss 1999 Video Study. <i>Educational Studies in Mathematics</i> . Vol.60, No.2, pp.199-215. |
| 2143 | D | |
| 2144A | A1 | Givvin, K B, Hiebert, J, Jacobs, J K, Hollingsworth, H & Gallimore, R (2005) Are There National Patterns of Teaching? Evidence from the TIMSS 1999 Video Study. <i>Comparative Education Review</i> . Vol.49, No.3, |

| # | C | Reference |
|-------|----|---|
| | | pp.311-343. |
| 2145A | A1 | Birenbaum, M, Tatsuoka, C & Tao, X (2005) Large-scale diagnostic assessment: comparison of eighth graders' mathematics performance in the United States, Singapore and Israel. <i>Assessment in Education: Principles, Policy & Practice</i> . Vol.12, No.2, pp.167-181. |
| 2146A | A2 | Hiebert, J, Stigler, J W, Jacobs, J K, Givvin, K B, Garnier, H, Smith, M, Hollingsworth, H, Manaster, A, Wearne, D & Gallimore, R (2005) Mathematics Teaching in the United States Today (And Tomorrow): Results from the TIMSS 1999 Video Study. <i>Educational Evaluation and Policy Analysis</i> . Vol.27, No.2, pp.111-132. |
| 2147A | NA | House, J D (2005) Classroom Instruction and Science Achievement in Japan, Hong Kong, and Chinese Taipei: Results from the TIMSS 1999 Assessment. <i>International Journal of Instructional Media</i> . Vol.32, No.3, pp.295- |
| 2148C | C | Wu, M (2005) The role of plausible values in large-scale surveys. <i>Studies in Educational Evaluation</i> . Vol.31, No.2/3, pp.114-128. |
| 2149A | A2 | Ferrini-Mundy, J & Schmidt, W H (2005) International Comparative Studies in Mathematics Education: Opportunities for Collaboration and Challenges for Researchers. <i>Journal for Research in Mathematics Education</i> . Vol.36, No.3, pp.164-175. |
| 2150 | D | |
| 2151A | A2 | Liu, X & Lesniak, K M (2005) Students' Progression of Understanding the Matter Concept from Elementary to High School. <i>Science Education</i> . Vol.89, No.3, pp.433-450. |
| 2152 | N | Howie, S & Plomp, T (2005) TIMSS-Mathematics Findings From National and International Perspectives: In search of explanations. <i>Educational Research & Evaluation</i> . Vol.11, No.2, pp.101-106. |
| 2153A | A2 | Van den Broeck, A, Opdenakker, M-C & Van Damme, J (2005) The Effects of Student Characteristics on Mathematics Achievement in Flemish TIMSS 1999 Data. <i>Educational Research & Evaluation</i> . Vol.11, No.2, pp.107-121. |
| 2154A | A2 | Howie, S J (2005) Contextual Factors at the School and Classroom Level Related to Pupils' Performance in Mathematics in South Africa. <i>Educational Research & Evaluation</i> . Vol.11, No.2, pp.123-140. |
| 2155B | B | Vos, P & Kuiper, W (2005) Trends (1995?-2000) in the TIMSS Mathematics Performance Assessment in The Netherlands Datations Exploratory Multigroup Approach Using Functional Data Analysis. <i>Educational Research & Evaluation</i> . Vol.11, No.2, pp.141-154. |
| 2156A | A1 | O'Dwyer, L M (2005) Examining the Variability of Mathematics Performance and its Correlates Using Data From TIMSS '95 and TIMSS '99. <i>Educational Research & Evaluation</i> . Vol.11, No.2, pp.155-177. |
| 2157A | A1 | Ce, S (2005) How American Middle Schools Differ From Schools of Five Asian Countries: Based on cross-national data from TIMSS 1999. <i>Educational Research & Evaluation</i> . Vol.11, No.2, pp.179-199. |
| 2158A | A2 | Vos, P & Bos, K (2005) The Mathematics Curriculum in The Netherlands: Measuring curricular alignment using TIMSS-99. <i>Educational Research & Evaluation</i> . Vol.11, No.2, pp.201-219. |
| 2159B | B | Crossley, M, Chisholm, L & Holmes, K (2005) Educational change and evaluation in Eastern and Southern Africa. <i>Compare: A Journal of Comparative Education</i> . Vol.35, No.1, pp.1-6. |
| 2160B | B | Reddy, V (2005) Cross-national achievement studies: learning from South Africa's participation in the Trends in International Mathematics and Science Study (TIMSS). <i>Compare: A Journal of Comparative Education</i> . Vol.35, No.1, pp.63-77. |
| 2161B | B | Sfard, A (2005) What Could be More Practical than Good Research? <i>Educational Studies in Mathematics</i> . Vol.58, No.3, pp.393-413. |
| 2162A | NA | House, J D (2005) Motivational Qualities of Instructional Strategies and Computer Use for Mathematics Teaching in Japan and the United States: Results from the TIMSS 1999 Assessment. <i>International Journal of Instructional Media</i> . Vol.32, No.1, pp.89-104. |
| 2163C | C | Ercikan, K & Kim, K (2005) Examining the Construct Comparability of the English and French Versions of TIMSS. <i>International Journal of Testing</i> . Vol.5, No.1, pp.23-35. |
| 2164C | C | Rupp, A A (2005) Quantifying Subpopulation Differences for a Lack of Invariance Using Complex Examinee Profiles: An Exploratory Multigroup Approach Using Functional Data Analysis. <i>Educational Research & Evaluation</i> . Vol.11, No.1, pp.71-97. |
| 2165 | N | Bybee, R W & Kennedy, D (2005) Math and Science Achievement. <i>Science</i> . Vol.307, No.5709, pp.481. |
| 2166 | NA | Pratt, H (2005) Where Are We Now? <i>Science Teacher</i> . Vol.72, No.1, pp.10-11. |
| 2167C | C | Hastedt, D & Sibberns, H (2005) Differences between Multiple Choice Items and Constructed Response Items in the IEA TIMSS Surveys. <i>Studies in Educational Evaluation</i> . Vol.31, No.2/3, pp.145-161. |
| 2168 | D | |
| 2169 | N | Bhattacharjee, Y (2004) Singapore Leads, U.S. Lags in Science, Math Student Achievement. <i>Science</i> . Vol.306, No.5705, p.2173. |
| 2170 | N | Grimm, D (2004) Hong Kong, Finland Students Top High School Test of Applied Skills. <i>Science</i> . Vol.306, No.5703, p.1877. |
| 2171A | A1 | Tatsuoka, K K, Corter, J E & Tatsuoka, C (2004) Patterns of Diagnosed Mathematical Content and Process Skills in TIMSS-R across a Sample of 20 Countries. <i>American Educational Research Journal</i> . Vol.41, No.4, pp.901-926. |

| # | C | Reference |
|-------|----|---|
| 2172A | NA | House, J D (2004) Cognitive-Motivational Characteristics and Science Achievement of Adolescent Students: Results from the TIMSS 1995 and TIMSS 1999 Assessments. <i>International Journal of Instructional Media</i> . Vol.31, No.4, pp.411-424. |
| 2173A | A2 | Sofroniou, N & Kellaghan, T (2004) The Utility of Third International Mathematics and Science Study Scales in Predicting Students' State Examination Performance. <i>Journal of Educational Measurement</i> . Vol.41, No.4, pp.311-329. |
| 2174A | A1 | Xin, T, Xu, Z & Tatsuoka, K (2004) Linkage between teacher quality, student achievement, and cognitive skills: A rule-space model. <i>Studies in Educational Evaluation</i> . Vol.30, No.4, pp.205-223. |
| 2175A | A1 | Papanastasiou, C & Papanastasiou, E C (2004) Major Influences on Attitudes Toward Science. <i>Educational Research & Evaluation</i> . Vol.10, No.3, pp.239-257. |
| 2176A | NA | House, J D (2004) The Effects of Homework Activities and Teaching Strategies for New Mathematics Topics on Achievement of Adolescent Students in Japan: Results from the TIMSS 1999 Assessment. <i>International Journal of Instructional Media</i> . Vol.31, No.2, pp.199-210. |
| 2177B | NA | Thompson, T D & Preston, R V (2004) Measurement in the Middle Grades: Insights from NAEP and TIMSS. <i>Mathematics Teaching in the Middle School</i> . Vol.9, No.9, pp.514-519. |
| 2178 | D | |
| 2179A | A1 | Papanastasiou, E C & Zembylas, M (2004) Differential effects of science attitudes and science achievement in Australia, Cyprus, and the USA. <i>International Journal of Science Education</i> . Vol.26, No.3, pp.259-280. |
| 2180 | M | Woodward, J & Ono, Y (2004) Mathematics and academic diversity in Japan. <i>Journal Of Learning Disabilities</i> . Vol.37, No.1, pp.74-82. |
| 2181A | NA | House, J D (2003) Instructional Activities and Interest in Science Learning for Adolescents in Japan and the United States: Findings from the Third International Mathematics and Science Study (TIMSS). <i>International Journal of Instructional Media</i> . Vol.30, No.4, pp.429-442. |
| 2182 | M | Schmidt, W H (2003) The Quest for a Coherent School Science Curriculum: The Need for an Organizing Principle. <i>Review of Policy Research</i> . Vol.20, No.4, pp.569-584. |
| 2183A | A1 | Hanushek, E A & Luque, J A (2003) Efficiency and equity in schools around the world. <i>Economics of Education Review</i> . Vol.22, No.5, pp.481- |
| 2184 | NA | Papanastasiou, E C, Zembylas, M & Vrasidas, C (2003) Can Computer Use Hurt Science Achievement? The USA Results from PISA. <i>Journal of Science Education and Technology</i> . Vol.12, No.3, pp.325-332. |
| 2185A | A1 | Vlaardingerbroek, B & Neil Taylor, T G (2003) Teacher education variables as correlates of primary science ratings in thirteen TIMSS systems. <i>International Journal of Educational Development</i> . Vol.23, No.4, pp.429- |
| 2186A | NA | House, J D (2003) Self-Belief and Science and Mathematics Achievement of Adolescent Students in Hong Kong: Findings from the Third International Mathematics and Science Study (TIMSS). <i>International Journal of Instructional Media</i> . Vol.30, No.2, pp.195- |
| 2187B | B | LeTendre, G K, Hofer, B K & Shimizu, H (2003) What Is Tracking? Cultural Expectations in the United States, Germany, and Japan. <i>American Educational Research Journal</i> . Vol.40, No.1, pp.43-89. |
| 2188 | NA | Schreiber, J B & Chambers, E A (2003) American High School Seniors' Mathematics Literacy Achievement. <i>North American Journal of Psychology</i> . Vol.5, No.1, pp.15-29. |
| 2189A | A1 | Wang, J & Zhu, C (2003) An In-depth Analysis of Achievement Gaps Between Seventh and Eighth Grades in the TIMSS Database. <i>School Science & Mathematics</i> . Vol.103, No.4, pp.186- |
| 2190A | NA | House, J D (2003) The Motivational Effects of Specific Instructional Strategies and Computer Use for Mathematics Learning in Japan: Findings from the Third International Mathematics and Science Study (TIMSS). <i>International Journal of Instructional Media</i> . Vol.30, No.1, pp.77- |
| 2191A | A2 | Patterson, M, Perry, E & Decker, C (2003) Factors Associated with High School Mathematics Performance in the United States. <i>Studies in Educational Evaluation</i> . Vol.29, No.2, pp.91-108. |
| 2192A | A1 | Akiba, M, LeTendre, G K, Baker, D P & Goesling, B (2002) Student Victimization: National and School System Effects on School Violence in 37 Nations. <i>American Educational Research Journal</i> . Vol.39, No.4, pp.829-853. |
| 2193A | NA | House, J D (2002) The Motivational Effects of Specific Teaching Activities and Computer Use for Science Learning: Findings from the Third International Mathematics and Science Study (TIMSS). <i>International Journal of Instructional Media</i> . Vol.29, No.4, pp.423- |
| 2194A | A2 | Howie, S & Plomp, T (2002) Mathematical literacy of school leaving pupils in South Africa. <i>International Journal of Educational Development</i> . Vol.22, No.6, pp.603- |
| 2195A | A2 | House, J D (2002) Instructional Practices and Mathematics Achievement of Adolescent Students in Chinese Taipei: Results from the TIMSS 1999 Assessment. <i>Child Study Journal</i> . Vol.32, No.3, pp.157- |
| 2196 | N | Weischadle, D E (2002) Extended Learning Opportunities: Some Lessons from the Field. <i>Education</i> . Vol.123, No.1, p.73. |
| 2197B | B | Baker, D P, Goesling, B & LeTendre, G K (2002) Socioeconomic Status, School Quality, and National Economic Development: A Cross-National Analysis of the "Heyneman-Loxley Effect" on Mathematics and Science Achievement. <i>Comparative Education Review</i> . Vol.46, No.3, pp.291-312. |
| 2198A | A1 | Schiller, K S, Khmelkov, V T & Wang, X-Q (2002) Economic Development and the Effects of Family |

| # | C | Reference |
|-------|----|---|
| | | Characteristics on Mathematics Achievement. <i>Journal of Marriage & Family</i> . Vol.64, No.3, pp.730-742. |
| 2199A | A1 | Shen, C (2002) Revisiting the Relationship Between Students' Achievement and their Self-perceptions: a cross-national analysis based on TIMSS 1999 data. <i>Assessment in Education: Principles, Policy & Practice</i> . Vol.9, No.2, pp.161-184. |
| 2200 | N | Bybee, R W (2002) Biology Education in the United States: The Unfinished Century. <i>BioScience</i> . Vol.52, No.7, p.560. |
| 2201A | A2 | Schreiber, J B (2002) Institutional and Student Factors and Their Influence on Advanced Mathematics Achievement. <i>The Journal of Educational Research</i> . Vol.95, No.5, pp.274-286. |
| 2202 | N | Papanastasiou, C (2002) Achievements in Mathematics and Science in an International Context. <i>Educational Research & Evaluation</i> . Vol.8, No.1, pp.3-12. |
| 2203A | A1 | Cogan, L S & Schmidt, W H (2002) "Culture Shock" – Eighth-Grade Mathematics From an International Perspective. <i>Educational Research & Evaluation</i> . Vol.8, No.1, pp.13-39. |
| 2204A | A1 | Kelly, D L (2002) The TIMSS 1995 International Benchmarks of Mathematics and Science Achievement: Profiles of World Class Performance at Fourth and Eighth Grades. <i>Educational Research & Evaluation</i> . Vol.8, No.1, pp.41-54. |
| 2205A | A2 | Weng, P (2002) Mathematics Teaching and Learning in Relation to the Danish Results in TIMSS Population 2. <i>Educational Research & Evaluation</i> . Vol.8, No.1, pp.41-54. |
| 2206A | A1 | Leung, F K S (2002) Behind the High Achievement of East Asian Students. <i>Educational Research & Evaluation</i> . Vol.8, No.1, pp.87-108. |
| 2207B | B | Vári, P, Tuska, A & Krolopp, J (2002) Change of Emphasis in the Mathematics Assessment in Hungary. <i>Educational Research & Evaluation</i> . Vol.8, No.1, pp.109-127. |
| 2208A | A1 | Papanastasiou, E (2002) Factors that Differentiate Mathematics Students in Cyprus, Hong Kong, and the USA. <i>Educational Research & Evaluation</i> . Vol.8, No.1, pp.129-146. |
| 2209A | A2 | Lattery, M J, Lemberger, J & Herzog, B (2002) Impact of the Science and Technology for Children Curriculum in the Oshkosh Area School District. <i>School Science & Mathematics</i> . Vol.102, No.3, pp.124. |
| 2210 | N | Martinez, J G R (2002) Building Conceptual Bridges from Arithmetic to Algebra. <i>Mathematics Teaching in the Middle School</i> . Vol.7, No.6, p.326. |
| 2211A | NA | House, J D (2002) The Use of Computers in a Mathematical Lesson in Japan: A Case Analysis from the TIMSS Videotape Classroom Study. <i>International Journal of Instructional Media</i> . Vol.29, No.1, pp.113-124. |
| 2212C | C | Ercikan, K (2002) Disentangling Sources of Differential Item Functioning in Multilanguage Assessments. <i>International Journal of Testing</i> . Vol.2, No.3/4, pp.199-215. |
| 2213B | B | Lawson, A, Bordignon, C & Nagy, P (2002) Matching the Grade 8 TIMSS Item Pool to the Ontario Curriculum. <i>Studies in Educational Evaluation</i> . Vol.28, No.1, pp.87-102. |
| 2214A | A2 | Papanastasiou, C (2002) School, Teaching and Family Influence on Student Attitudes toward Science, Based on TIMSS Data for Cyprus. <i>Studies in Educational Evaluation</i> . Vol.28, No.1, pp.71-86. |
| 2215A | A2 | Papanastasiou, C (2002) TIMSS Study in Cyprus: Patterns of Achievements in Mathematics and Science. <i>Studies in Educational Evaluation</i> . Vol.28, No.3, pp.223-233. |
| 2216A | A2 | Cogan, L S, Schmidt, W H & Wiley, D E (2001) Who Takes What Math and in Which Track? Using TIMSS to Characterize U.S. Students' Eighth-Grade Mathematics Learning Opportunities. <i>Educational Evaluation and Policy Analysis</i> . Vol.23, No.4, pp.323-341. |
| 2217A | A1 | Pong, S-I & Pallas, A (2001) Class Size and Eighth-Grade Math Achievement in the United States and Abroad. <i>Educational Evaluation and Policy Analysis</i> . Vol.23, No.3, pp.251-273. |
| 2218B | B | Lindquist, M M (2001) NAEP, TIMSS, and PSSM: Entangled Influences. <i>School Science & Mathematics</i> . Vol.101, No.6, pp.286- |
| 2219B | B | LeTendre, G K, Baker, D P, Akiba, M, Goesling, B & Wiseman, A (2001) Teachers' Work: Institutional Isomorphism and Cultural Variation in the U.S., Germany, and Japan. <i>Educational Researcher</i> . Vol.30, No.6, pp.3-15. |
| 2220C | C | Wang, J (2001) TIMSS Primary and Middle School Data: Some Technical Concerns. <i>Educational Researcher</i> . Vol.30, No.6, pp.17-21. |
| 2221B | B | Phelps, R P (2001) Benchmarking to the World's Best in Mathematics: Quality Control in Curriculum and Instruction Among the Top Performers in the TIMSS. <i>Evaluation Review</i> . Vol.25, No.4, pp.391-349. |
| 2222A | NA | Shen, C (2001) Social Values Associated with Cross-national Differences in Mathematics and Science Achievement: a cross-national analysis. <i>Assessment in Education: Principles, Policy & Practice</i> . Vol.8, No.2, pp.193-223. |
| 2223 | N | ScienceScope. <i>Science</i> . Vol.292, No.5515, p.187. |
| 2224A | A2 | O'Leary, M (2001) The Effects of Age-based and Grade-based Sampling on the Relative Standing of Countries in International Comparative Studies of Student Achievement. <i>British Educational Research Journal</i> . Vol.27, No.2, pp.187-200. |
| 2225A | A1 | Baker, D P, Akiba, M, LeTendre, G K & Wiseman, A W (2001) Worldwide Shadow Education: Outside-School Learning, Institutional Quality of Schooling, and Cross-National Mathematics Achievement. <i>Educational Evaluation and Policy Analysis</i> . Vol.23, No.1, pp.1-17. |

| # | C | Reference |
|-------|----|--|
| 2226C | NA | Romance, N R & Vitale, M R (2001) Implementing an in-depth expanded science model in elementary schools: Multi-year findings, research issues, and policy implications. <i>International Journal of Science Education</i> . Vol.23, No.4, pp.373-404. |
| 2227A | A1 | Bielinski, J & Davison, M L (2001) A Sex Difference by Item Difficulty Interaction in Multiple-Choice Mathematics Items Administered to National Probability Samples. <i>Journal of Educational Measurement</i> . Vol.38, No.1, pp.51-77. |
| 2228A | NA | House, J D (2001) Relationships Between Instructional Activities and Mathematics Achievement of Adolescent Students in Japan: Findings from the Third International Mathematics and Science Study (TIMSS). <i>International Journal of Instructional Media</i> . Vol.28, No.1, pp.93- |
| 2229A | A2 | Casey, M B, Nuttall, R L & Pezaris, E (2001) Spatial-Mechanical Reasoning Skills versus Mathematics Self-Confidence as Mediators of Gender Differences on Mathematics Subtests Using Cross-National Gender-Based Items. <i>Journal for Research in Mathematics Education</i> . Vol.32, No.1, pp.28-57. |
| 2230A | A2 | Koller, O (2001) Mathematical World Views and Achievement in Advanced Mathematics in Germany: Findings from TIMSS Population 3. <i>Studies in Educational Evaluation</i> . Vol.27, No.1, pp.65-78. |
| 2231 | N | Kuiper, W, Bos, K & Plomp, T (2001) Mathematics and Science Literacy of Final-Year Secondary School Students: Dutch Results in an International Context. <i>Studies in Educational Evaluation</i> . Vol.27, No.1, pp.27-46. |
| 2232A | A2 | Howie, S L & Pietersen, J J (2001) Mathematics Literacy of Final Year Students: South African Realities. <i>Studies in Educational Evaluation</i> . Vol.27, No.1, pp.7-25. |
| 2233 | N | Kuiper, W & Plomp, T (2001) National and Cross-National Perspective on TIMSS Populations 2 and 3 Findings. <i>Studies in Educational Evaluation</i> . Vol.27, No.1, pp.1-25. |
| 2234A | A2 | Ramseier, E (2001) Scientific Literacy of Upper-Secondary Students: A Swiss Perspective. <i>Studies in Educational Evaluation</i> . Vol.27, No.1, pp.47-64. |
| 2235A | A1 | Cheng, S K & Seng, Q K (2001) Sex Differences in TIMSS Mathematics Achievement of Four Asian Nations. <i>Studies in Educational Evaluation</i> . Vol.27, No.4, pp.331-340. |
| 2236A | A2 | Bos, K, Kuiper, W & Plomp, T (2001) TIMSS Results of Dutch Grade 8 Students in International Perspective: Performance Assessment and Written Test. <i>Studies in Educational Evaluation</i> . Vol.27, No.1, pp.79-94. |
| 2237 | N | Holden, C (2000) Asia Stays on Top, U.S. in Middle In New Global Rankings. <i>Science</i> . Vol.290, No.5498, p.1866. |
| 2238A | A2 | Wilkins, J L M (2000) Preparing for the 21st Century: The Status of Quantitative Literacy in the United States. <i>School Science & Mathematics</i> . Vol.100, No.8, pp.405- |
| 2239A | A1 | Post, D & Pong, S-I (2000) Employment during Middle School: The Effects on Academic Achievement in the U. S. and Abroad. <i>Educational Evaluation and Policy Analysis</i> . Vol.22, No.3, pp.273-298. |
| 2240 | D | |
| 2241A | NA | House, J D (2000) Relationships Between Instructional Activities and Science Achievement of Adolescent Students in Hong Kong: Findings from the Third International Mathematics and Science Study (TIMSS). <i>International Journal of Instructional Media</i> . Vol.27, No.3, p.275- |
| 2242A | A1 | Shen, C & Pedulla, J J (2000) The Relationship between Students' Achievement and their Self-perception of Competence and Rigour of Mathematics and Science: a cross-national analysis. <i>Assessment in Education: Principles, Policy & Practice</i> . Vol.7, No.2, pp.237-253. |
| 2243C | C | Wang, J (2000) Relevance of the hierarchical linear model to TIMSS data analyses. <i>Education</i> . Vol.120, No.4, pp.787-781. |
| 2244B | B | Macnab, D (2000) Raising standards in mathematics education: values, vision, and TIMSS. <i>Educational Studies in Mathematics</i> . Vol.42, No.1, pp61-80. |
| 2245C | C | Bracey, G W (2000) The TIMSS "Final Year" Study and Report: A Critique. <i>Educational Researcher</i> . Vol.29, No.4, pp.4-10. |
| 2246A | NA | House, J D (2000) Student self-beliefs and science achievement in Ireland: Findings from the Third International Mathematics and Science Study (TIMMS). <i>International Journal of Instructional Media</i> . Vol.27, No.1, pp.107- |
| 2247A | A1 | Tarr, J E, Uekawa, K, Mittag, K C & Lennex, L (2000) A Comparison of Calculator Use in Eighth-Grade Mathematics Classrooms in the United States, Japan, and Portugal: Results From the Third International Mathematics and Science Study. <i>School Science & Mathematics</i> . Vol.100, No.3, pp.139- |
| 2248B | B | Lassibille, G & Gomez, L N (2000) Organisation and Efficiency of Education Systems: some empirical findings. <i>Comparative Education</i> . Vol.36, No.1, pp.7-19. |
| 2249B | NA | Menon, R (2000) Should the United States emulate Singapore's education system to achieve Singapore's success in the TIMSS? <i>Mathematics Teaching in the Middle School</i> . Vol.5, No.6, pp.345- |
| 2250 | D | |
| 2251B | B | Lee, O & Paik, S-h (2000) Conceptions of Science Achievement in Major Reform Documents. <i>School Science & Mathematics</i> . Vol.100, No.1, pp.16- |
| 2252 | N | Forgione, P & Nohara, D (2000) Secondary performance. <i>Science Teacher</i> . Vol.67, No.1, pp.60- |
| 2253 | N | Rakow, S J (2000) NSTA's response to the TIMSS. <i>Science Teacher</i> . Vol.67, No.1, pp.61- |

| # | C | Reference |
|-------|----|---|
| 2254C | C | Papanastasiou, C (2000) Effects of Attitudes and Beliefs on Mathematics Achievement. <i>Studies in Educational Evaluation</i> . Vol.26, No.1, pp.27-42. |
| 2255 | N | Papanastasiou, C (2000) Internal and External Factors Affecting Achievement in Mathematics: Some Findings from TIMSS. <i>Studies in Educational Evaluation</i> . Vol.26, No.1, pp.1-7. |
| 2256A | A2 | Lokan, J & Greenwood, L (2000) Mathematics Achievement at Lower Secondary Level in Australia. <i>Studies in Educational Evaluation</i> . Vol.26, No.1, pp.9-26. |
| 2257A | A2 | Howie, S J, Marsh, T A & Allummoottil, J (2000) Middle School Students' Performance in Mathematics in the Third International Mathematics and Science Study: South African Realities. <i>Studies in Educational Evaluation</i> . Vol.26, No.1, pp.61-77. |
| 2258A | A2 | Wester, A & Henriksson, W (2000) The Interaction between Item Format and Gender Differences in Mathematics Performance Based on TIMSS Data. <i>Studies in Educational Evaluation</i> . Vol.26, No.1, pp.79-90. |
| 2259B | B | Kuiper, W, Bos, K & Plomp T (2000) The TIMSS National Option Mathematics Test. <i>Studies in Educational Evaluation</i> . Vol.26, No.1, pp.43-60. |
| 2260A | A2 | Zuzovsky, R & Tamir, P (1999) Growth patterns in students' ability to supply scientific explanations: findings from the Third International Mathematics and Science Study in Israel. <i>International Journal of Science Education</i> . Vol.21, No.10, pp.1101-1121. |
| 2261A | A2 | Moreno, N P (1999) K-12 science education reform-a primer for scientists. <i>BioScience</i> . Vol.49, No.7, pp.569- |
| 2262 | N | Kuiper, W & Plomp, T (1999) Special Issue: TIMSS in a Western European Context. <i>Educational Research & Evaluation</i> . Vol.5, No.2, p.77. |
| 2263 | N | Kuiper, W & Plomp, T (1999) TIMSS in a Western European Context. <i>Educational Research & Evaluation</i> . Vol.5, No.2, p.79-84. |
| 2264A | A2 | Kuiper, W, Bos, K & Plomp, T (1999) Mathematics Achievement in the Netherlands and Appropriateness of the TIMSS Mathematics Test. <i>Educational Research & Evaluation</i> . Vol.5, No.2, pp.105- |
| 2265A | A2 | Ramseier, E (1999) Task Difficulty and Curricular Priorities in Science: Analysis of Typical Features of the Swiss Performance in TIMSS. <i>Educational Research & Evaluation</i> . Vol.5, No.2, pp.105- |
| 2266A | A2 | Adolfsson, L & Henriksson, W (1999) Different Methods – Different Results: How the Swedish Results in Mathematics Vary Depending on Methodological Approach. <i>Educational Research & Evaluation</i> . Vol.5, No.2, pp.127-138. |
| 2267A | A2 | Huber, M (1999) Co-ordination Within Schools, Commitment of Teachers and Students and Student Achievement. <i>Educational Research & Evaluation</i> . Vol.5, No.2, pp.139- |
| 2268A | A1 | Bos, K & Kuiper, W (1999) Modelling TIMSS Data in a European Comparative Perspective: Exploring Influencing Factors on Achievement in Mathematics in Grade 8. <i>Educational Research & Evaluation</i> . Vol.5, No.2, pp.157- |
| 2269A | A2 | Köller, O, Baumert, J, Clausen, M & Hosenfeld, I (1999) Predicting Mathematics Achievement of Eighth Grade Students in Germany: An Application of Parts of the Model of Educational Productivity to the TIMSS Data. <i>Educational Research & Evaluation</i> . Vol.5, No.2, pp.180- |
| 2270A | A2 | Keys, W (1999) What Can Mathematics Educators in England Learn from TIMSS? <i>Educational Research & Evaluation</i> . Vol.5, No.2, pp.195-213. |
| 2271A | A2 | Knuver, A (1999) Mathematics and Science Performance of Primary School Students in the Netherlands. <i>Educational Research & Evaluation</i> . Vol.5, No.2, pp.214- |
| 2272C | C | Cheng, Y C & Cheung, W M (1999) Lessons from TIMSS in Europe: An Observation from Asia. <i>Educational Research & Evaluation</i> . Vol.5, No.2, pp.227-236. |
| 2273B | B | Beaton, A A E (1999) Comments on Secondary Analysis of the TIMSS Data. <i>Educational Research & Evaluation</i> . Vol.5, No.2, pp.237-239. |
| 2274B | B | Schmidt, W H & Prawat, R S (1999) What Does the Third International Mathematics and Science Study Tell Us about Where to Draw the Line in the Top-Down versus Bottom-Up Debate? <i>Educational Evaluation and Policy Analysis</i> . Vol.21, No.1, pp.85-91. |
| 2275B | B | Okhee, L (1999) Equity Implications Based on the Conceptions of Science Achievement in Major Reform Documents. <i>Review of Educational Research</i> . Vol.69, No.1, pp.83-115. |
| 2276C | C | Olson, S (1999) Videotapes Expose Classroom Faults. <i>Science</i> . Vol.283, No.5408, pp.1616-1617. |
| 2277B | B | Setinc, M (1999) International Comparative Studies of Science Achievement from the Time Perspective: the Third International Mathematics and Science Study versus the International Assessment of Educational Progress. <i>Assessment in Education: Principles, Policy & Practice</i> . Vol.6, No.1, pp.43- |
| 2278C | C | Lokan, A R & Doig, B (1999) Broadening Assessment, Improving Fairness? Some examples from school science. <i>Assessment in Education: Principles, Policy & Practice</i> . Vol.6, No.1, pp.83- |
| 2279 | NA | McMahon, M M (1999) Rededicating ourselves to reform. <i>Science Teacher</i> . Vol.66, No.1, pp.8- |
| 2280A | A2 | Neuschatz, M (1999) What can the TIMSS teach us? <i>Science Teacher</i> . Vol.66, No.1, pp.23- |
| 2281B | NA | Pauls, J & Young, D B (1999) Laboratory for learning? <i>Science Teacher</i> . Vol.66, No.1, p.27- |
| 2282 | NA | Rakow, S J (1999) NSTA's response to the TIMSS. (cover story). <i>Science Teacher</i> . Vol.66, No.1, pp.30- |
| 2283 | NA | Holliday, W G (1999) Questioning the TIMSS. <i>Science Teacher</i> . Vol.66, No.1, pp.34- |

| # | C | Reference |
|-------|----|---|
| 2284 | M | Peterman, L & Linder-Scholer, B (1999) The value of an A. <i>Science Teacher</i> . Vol.66, No.1, pp.42- |
| 2285 | N | Hurd, P D (1999) Comparing student achievement. <i>Science Teacher</i> . Vol.66, No.1, pp.46- |
| 2286C | C | Garden, R A (1999) Development of TIMSS Performance Assessment Tasks. <i>Studies in Educational Evaluation</i> . Vol.25, No.3, pp.217-241. |
| 2287C | C | Lokan, J (1999) Equity Issues in Testing: The Case of TIMSS Performance Assessment. <i>Studies in Educational Evaluation</i> . Vol.25, No.3, pp.297-314. |
| 2288C | C | Zuzovsky, R (1999) Performance Assessment in Science: Lessons from the Practical Assessment of 4th Grade Students in Israel. <i>Studies in Educational Evaluation</i> . Vol.25, No.3, pp.195-216. |
| 2289A | A2 | Paleckova, J (1999) Performance Assessment in the Czech Republic. <i>Studies in Educational Evaluation</i> . Vol.25, No.3, pp.265-268. |
| 2290C | C | Harmon, M (1999) Performance Assessment in the Third International Mathematics and Science Study--An International Perspective. <i>Studies in Educational Evaluation</i> . Vol.25, No.3, pp.243-262. |
| 2291A | A2 | Tamir, P & Zuzovsky, R (1999) Performance of Israeli 4th and 8th Grade Students on Science and Combined Science and Math Performance Assessment Tasks. <i>Studies in Educational Evaluation</i> . Vol.25, No.3, pp.283-286. |
| 2292C | C | Zuzovsky, R (1999) Problematic Aspects of the Scoring System of the TIMSS Practical Performance Assessment: Some Examples. <i>Studies in Educational Evaluation</i> . Vol.25, No.3, pp.315-323. |
| 2293B | B | Bos, K, Kuiper, W A & Plomp, T (1999) Student Performance and Curricular Appropriateness in the Netherlands. <i>Studies in Educational Evaluation</i> . Vol.25, No.3, pp.269-276. |
| 2294A | A1 | Philippou, G N & Christou, C (1999) Teachers' Conceptions of Mathematics and Students' Achievement: A Cross-Cultural Study Based on Results from TIMSS. <i>Studies in Educational Evaluation</i> . Vol.25, No.4, pp.379-398. |
| 2295A | A2 | Jakwerth, P (1999) TIMSS Performance Assessment Results: United States. <i>Studies in Educational Evaluation</i> . Vol.25, No.3, pp.277-281. |
| 2296A | A2 | Harris, S (1999) TIMSS Performance Assessment: The Strengths and Weaknesses of Students in England. <i>Studies in Educational Evaluation</i> . Vol.25, No.3, pp.287-295. |
| 2297A | A2 | Hosenfeld, I, Koller, O & Baumert, J (1999) Why Sex Differences in Mathematics Achievement Disappear in German Secondary Schools: A Reanalysis of the German TIMSS-Data. <i>Studies in Educational Evaluation</i> . Vol.25, No.2, pp.143-161. |
| 2298B | B | Schmidt, W H & McKnight, C C (1998) What Can We Really Learn from TIMSS? <i>Science</i> . Vol.282, No.5395, pp.1830-1831. |
| 2299 | NA | Friel, S N (1998) Exploring How One Problem Contributes to Student Learning. <i>Mathematics Teaching in the Middle School</i> . Vol.4, No.2, p.100. |
| 2300 | NA | Reys, R E (1998) Computation versus Number Sense. <i>Mathematics Teaching in the Middle School</i> . Vol.4, No.2, p.110. |
| 2301 | NA | Zucker, A A (1998) It's time to talk about textbooks. <i>Mathematics Teaching in the Middle School</i> . Vol.4, No.1, p.16. |
| 2302 | NA | Ferrini-Mundy, J (1998) Learning from the Math standards. <i>Science Teacher</i> . Vol.65, No.6, p.27. |
| 2303 | NA | Pratt, H (1998) Focusing on achievement. <i>Science Teacher</i> . Vol.65, No.6, p.30. |
| 2304 | N | Maienschein, J (1998) Scientific literacy. <i>Science</i> . Vol.281, No.5379, p.917. |
| 2305 | N | Rotberg, I C (1998) Interpretation of international test score comparisons. <i>Science</i> . Vol.280, No.5366, p.1030-1031. |
| 2306B | B | Wolf, R M (1998) National Standards: Do We Need Them? <i>Educational Researcher</i> . Vol.27, No.4, pp.22-25. |
| 2307 | N | Holden, C (1998) Rehashing TIMSS. <i>Science</i> . Vol.280, No.5362, p.385. |
| 2308 | N | Greenberg, D S (1998) Stumbling in science studies. <i>Lancet</i> . Vol.351, No.9109, p.1142. |
| 2309 | N | Wheeler, G (1998) The wake-up call we dare not ignore. <i>Science</i> . Vol.279, No.5357, p.1611. |
| 2310A | NA | McKnight, CC & Schmidt, W H (1998) Facing Facts in U.S. Science and Mathematics Education: Where We Stand, Where We Want to Go. <i>Journal of Science Education and Technology</i> . Vol.7, No.1, pp.57-76. |
| 2311 | N | Vogel, G (1998) Northern Europe tops in high school. <i>Science</i> . Vol.279, No.5355, p.1297. |
| 2312 | N | Vogel, G (1997) Students don't measure up to standards. <i>Science</i> . Vol.278, No.5339, p.794. |
| 2313 | N | Linn, M C (1997) Finding patterns in international assessments. <i>Science</i> . Vol.277, No.5333, p.1743. |
| 2314 | N | Vogel, G (1997) U.S. kids score well in primary grades. <i>Science</i> . Vol.276, No.5319, p.1642. |
| 2315A | A2 | Stedman, L C (1997) International Achievement Differences: An Assessment of a New Perspective. <i>Educational Researcher</i> . Vol.26, No.3, pp.4-15. |
| 2316 | N | Vogel, G (1996) Asia and Europe top in world, but reasons are hard to find. <i>Science</i> . Vol.274, No.5291, p.1296. |
| 2317C | C | Schmidt, W H & McKnight, C C (1995) Surveying Educational Opportunity in Mathematics and Science: An International Perspective. <i>Educational Evaluation and Policy Analysis</i> . Vol.17, No.3, pp.337-353. |
| 30001 | M | Pennington, J L, Obenchain, K M & Brock, C H (2014) Reading Informational Texts. <i>Reading Teacher</i> . Vol.67, No.7, pp.532-542. |
| 3002A | A1 | Isac, M M, Maslowski, R, Creemers, B & van der Werf, G (2014) The contribution of schooling to |

| # | C | Reference |
|-------|----|---|
| | | secondary-school students' citizenship outcomes across countries. <i>School Effectiveness & School Improvement</i> . Vol.25, No.1, pp.29-63. |
| 3003A | A1 | Lin, A R (2014) Examining Students' Perception of Classroom Openness as a Predictor of Civic Knowledge: A Cross-National Analysis of 38 Countries. <i>Applied Developmental Science</i> . Vo.18, No.1, pp.17-30. |
| 3004B | B | Guérin, L J F, van der Ploeg, P A & Sins, P H M (2013) Citizenship education: the feasibility of a participative approach. <i>Educational Research</i> . Vol.55, No.4, pp.427-440. |
| 3005A | A1 | Quintelier, E & Hooghe, M (2013) The relationship between political participation intentions of adolescents and a participatory democratic climate at school in 35 countries. <i>Oxford Review of Education</i> . Vol.39, No.5, pp.567-589. |
| 3006 | M | Maloyed, C L & Williams, J K (2013) Place-Based Civic Education and the Rural Leadership Crisis in Nebraska. <i>Great Plains Research</i> . Vol.23, No.2, pp.127-135. |
| 3007A | A1 | Elchardus, M, Franck, E, de Groof, S & Kavadias, D (2013) The Acceptance of the Multicultural Society Among Young People. A Comparative Analysis of the Effect of Market-Driven Versus Publicly Regulated Educational Systems. <i>European Sociological Review</i> . Vol.29, No.4, pp.767-779. |
| 3008A | A1 | Lauglo, J (2013) Do more knowledgeable adolescents have more rationally based civic attitudes? Analysis of 38 countries. <i>Educational Psychology</i> . Vol.33, No.3, pp.255-275. |
| 3009C | C | Schulz, W, Fraillon, J & Ainley, J (2013) Measuring young people's understanding of civics and citizenship in a cross-national study. <i>Educational Psychology</i> . Vol.33, No.3, pp.327-349. |
| 3010A | A1 | Barber, C, Fennely, K & Torney-Purta, J (2013) Nationalism and Support for Immigrants' Rights Among Adolescents in 25 Countries. <i>Applied Developmental Science</i> . Vol.17, No.2, pp.60-75. |
| 3011A | A1 | Hooghe, M & Dassonneville, R (2013) Voters and Candidates of the Future: The Intention of Electoral Participation among Adolescents in 22 European Countries. <i>Young</i> . Vol.21, No.1, pp.1-28. |
| 3012C | NA | Chow, J (2012) Establishing a general framework civic competency for European youth. <i>International Journal of Progressive Education</i> . Vol.8, No.3, pp.140-150. |
| 3013C | C | Chow, J & Kennedy, K (2012) Citizenship and Governance in the Asian Region: Insights from The International Civic and Citizenship Education Study. <i>Public Organization Review</i> . Vol.12, No.3, pp.299-311. |
| 3014 | M | Dassonneville, R, Quintelier, E, Hooghe, M & Claes, E (2012) The Relation Between Civic Education and Political Attitudes and Behavior: A Two-Year Panel Study Among Belgian Late Adolescents. <i>Applied Developmental Science</i> . Vol.16, No.3, pp.140-150. |
| 3015A | A1 | Hoskins, B, Janmaat, J G & Villalba, E (2012) Learning citizenship through social participation outside and inside school: an international, multilevel study of young people's learning of citizenship. <i>British Educational Research Journal</i> . Vol.38, No.3, pp.419-446. |
| 3016 | M | Aroge, S T (2012) Civic Education as a Panacea to Electoral Malpractices in Nigeria. <i>Business & Management Research</i> . Vol.1, No.1, pp.141-146. |
| 3017 | M | McCowan, T (2012) Opening spaces for citizenship in higher education: three initiatives in English universities. <i>Studies in Higher Education</i> . Vol.37, No.1, pp.51-67. |
| 3018A | A1 | Janmaat, J G (2012) The Effect of Classroom Diversity on Tolerance and Participation in England, Sweden and Germany. <i>Journal of Ethnic & Migration Studies</i> . Vol.38, No.1, pp.21-39. |
| 3019A | A1 | Torney-Purta, J B C (2011) Fostering young people's support for participatory human rights through their developmental niches. <i>The American Journal Of Orthopsychiatry</i> . Vol.81, No.4, pp.473-481. |
| 3020 | M | Yusuf, A R, Agbonna, S A, Jekayinfa, A A & Salu, A (2011) Effects of citizenship education component of social studies on civic literacy and attachment of upper basic students in Ilorin Metropolis, Nigeria. <i>African Journal of Political Science & International Relations</i> . Vol.5, No.9, pp.437-441. |
| 3021A | A1 | Isac, M M, Maslowski, R & van der Werf, G (2011) Effective civic education: an educational effectiveness model for explaining students' civic knowledge. <i>School Effectiveness & School Improvement</i> . Vol.22, No.3, pp.313-333. |
| 3022 | M | Winstead, L (2011) The Impact of NCLB and Accountability on Social Studies: Teacher Experiences and Perceptions about Teaching Social Studies. <i>Social Studies</i> . Vol.102, No.5, pp.221-227. |
| 3023A | A1 | Janmaat, J G (2011) Ability grouping, segregation and civic competences among adolescents. <i>International Sociology</i> . Vol.26, No.4, pp.455-482. |
| 3024 | M | Bromley, P (2011) Multiculturalism and human rights in civic education: the case of British Columbia, Canada. <i>Educational Research</i> . Vol.53, No.2, pp.151-164. |
| 3025A | A1 | Hoskins, B L, Barber, C, van Nijlen, D & Villalba, E (2011) Comparing Civic Competence among European Youth: Composite and Domain-Specific Indicators Using IEA Civic Education Study Data. <i>Comparative Education Review</i> . Vol.55, No.1, pp.82-110. |
| 3026B | B | Rutkowski, D & Engel, L C (2010) Soft Power and Hard Measures: large-scale assessment, citizenship and the European Union. <i>European Educational Research Journal</i> . Vol.9, No.3, pp.6- |
| 3027 | M | Feldmann, D (2010) The Creation of an Effective Pilot Course in High School Civics: The Kentucky Model. <i>Ohio Social Studies Review</i> . Vol.46, No.2, pp.29-35. |
| 3028 | M | Joshi, P & Marri, A R (2010) An Economics Methods Course?: Challenges of Teaching an Economics |

| # | C | Reference |
|-------|----|---|
| | | Education Methods Course for Secondary Social Studies Preservice Teachers. <i>Social Studies</i> . Vol.97, No.5, pp.197-202. |
| 3029A | A2 | Doğanay, A (2010) What does democracy mean to 14-year-old Turkish children? A comparison with results of the 1999 IEA Civic Education Study. <i>Research Papers in Education</i> . Vol.25, No.1, pp.51-71. |
| 3030 | D | |
| 3031B | B | Torney-Purta, J V (2009) International Psychological Research That Matters for Policy and Practice. <i>American Psychologist</i> . Vol.64, No.8, pp.825-837. |
| 3032 | D | |
| 3033 | D | |
| 3034 | M | Harris, V J (2009) We, Too, Sing America. <i>Reading Teacher</i> . Vol.62, No.5, pp.450-455. |
| 3035A | A1 | Torney-Purta, J, Wilkenfeld, B & Barber, C (2008) How Adolescents in 27 Countries Understand, Support, and Practice Human Rights. <i>Journal of Social Issues</i> . Vol.64, No.4, pp.857-880. |
| 3036A | A2 | Campbell, D E (2008) Voice in the Classroom: How an Open Classroom Climate Fosters Political Engagement among Adolescents. <i>Political Behavior</i> . Vol.30, No.4, pp.437-454. |
| 3037A | A1 | Alviar-Martin, T, Randall, J D, Usher, E L & Engelhard, G (2008) Teaching Civic Topics in Four Societies: Examining National Context and Teacher Confidence. <i>Journal of Educational Research</i> . Vol.101, No.3, pp.177-188. |
| 3038A | A1 | Kennedy, K J (2007) Student Constructions of 'Active Citizenship': What Does Participation Mean to Students? <i>British Journal of Educational Studies</i> . Vol.55, No.3, pp.304-324. |
| 3039A | A2 | Torney-Purta, J, Barber, C & Wilkenfeld, B (2007) Latino Adolescents' Civic Development in the United States: Research Results from the IEA Civic Education Study. <i>Journal of Youth & Adolescence</i> . Vol.36, No.2, pp.111-125. |
| 3040C | C | Buk-Berge, E (2006) Missed opportunities: the IEA's study of civic education and civic education in post-communist countries. <i>Comparative Education</i> . Vol.42, No.4, pp.533-548. |
| 3041 | D | |
| 3042 | M | Lesser, E J (2006) Constitution Day: Start the School Year with Civics. <i>Social Education</i> . Vol.70, No.5, pp.294-300. |
| 3043A | A2 | Kennedy, K J & Mellor, S (2006) Australian Students' Civic Attitudes as Indicators of Support for Social Capital: Learning outcomes for the future. <i>Educational Psychology</i> . Vol.26, No.2, pp.251-271. |
| 3044C | C | Yuk, F C (2006) Analysis of School Context Effects on Differential Item Functioning Using Hierarchical Generalized Linear Models. <i>International Journal of Testing</i> . Vol.6, No.1, pp.57-79. |
| 3045 | D | |
| 3046 | M | Ruget, V (2006) The renewal of civic education in France and in America: Comparative perspectives. <i>Social Science Journal</i> . Vol.43, No.1, pp.19-34. |
| 3047 | M | Beaumont, E (2005) The Challenge of Assessing Civic Engagement: What We Know and What We Still Need to Learn about Civic Education in College. <i>Journal of Public Affairs Education</i> . Vol.11, No.4, pp.287-303. |
| 3048 | M | Edwards, J D (2005) JNCL-NCLIS: National Language Activities and Policies 2005. <i>Learning Languages</i> . Vol.10, No.2, p.10. |
| 3049A | A2 | Ichilov, O (2005) Pride in One's Country and Citizenship Orientations in a Divided Society: The Case of Israeli Palestinian Arab and Orthodox and Non-Orthodox Jewish Israeli Youth. <i>Comparative Education Review</i> . Vol.49, No.1, pp.44- |
| 3050 | M | Branson, M S (2005) The connection between civic and economic education. <i>Teacher Librarian</i> . Vol.32, No.3, pp.26-28. |
| 3051 | M | Kubey, R (2004) Media Literacy and the Teaching of Civics and Social Studies at the Dawn of the 21st Century. <i>American Behavioral Scientist</i> . Vol.48, No.1, pp.69-77. |
| 3052C | C | Torney-Purta, J (2004) Adolescents' Political Socialization in Changing Contexts: An International Study in the Spirit of Nevitt Sanford. <i>Political Psychology</i> . Vol.25, No.3, pp.465-478. |
| 3053 | M | Murphy, T A (2004) Deliberative civic education and civil society: a consideration of ideals and actualities in democracy and communication education. <i>Communication Education</i> . Vol.53, No.1, pp.74-91. |
| 3054 | M | Murphy, J B (2004) Against Civic Schooling. <i>Social Philosophy & Policy</i> . Vol.21, No.1, pp.221-265. |
| 3055A | A2 | Mellor, S J & Kennedy, K (2003) Australian Students' democratic values and attitudes towards participation: indicators from the IEA civic education study. <i>International Journal of Educational Research</i> . Vol.39, No.6, pp.525-537. |
| 3056A | A2 | Papanastasiou, C & Koutselini, M (2003) Developmental model of democratic values and attitudes toward social actions. <i>International Journal of Educational Research</i> . Vol.39, No.6, pp.539-549. |
| 3057A | A2 | Losito, B & D'Ápice, A (2003) Democracy, citizenship, participation. The results of the second IEA civic education study in Italy. <i>International Journal of Educational Research</i> . Vol.39, No.6, pp.609-620. |
| 3058A | NA | Fjeldstad, D & Mikkelsen, R (2003) Strong democratic competence does not automatically lead to strong engagement and participation. <i>International Journal of Educational Research</i> . Vol.39, No.6, pp.621-632. |
| 3059A | A2 | Hahn, C L (2003) Democratic values and citizen action: a view from US ninth graders. <i>International Journal of Educational Research</i> . Vol.39, No.6, pp.633-642. |

| # | C | Reference |
|-------|----|---|
| 3060 | M | Colby, A, Ehrlich, T, Beaumont, E & Stephens, J (2003) Educating Undergraduates for Responsible Citizenship. <i>Change</i> . Vol.35, No.6, pp.40-48. |
| 3061B | B | Torney-Purta, J (2003) A European Perspective on the IEA Civic Education Study: an introduction. <i>European Educational Research Journal</i> . Vol.2, No.3, pp.366-369. |
| 3062C | C | Nikolova, R & Lehmann, R H (2003) On the Dimensionality of the Cognitive Test Used in the IEA Civic Education Study: analyses and implications. <i>European Educational Research Journal</i> . Vol.2, No.3, pp.370-383. |
| 3063A | A1 | Maiello, C, Oser, F & Biedermann, H (2003) Civic Knowledge, Civic Skills and Civic Engagement. <i>European Educational Research Journal</i> . Vol.2, No.3, pp.284-395. |
| 3064C | C | Husfeldt, V & Nikolova, R (2003) Students' Concepts of Democracy. <i>European Educational Research Journal</i> . Vol.2, No.3, pp.396-409. |
| 3065A | A1 | Menezes, I (2003) Participation Experiences and Civic Concepts, Attitudes and Engagement: implications for citizenship education projects. <i>European Educational Research Journal</i> . Vol.2, No.3, pp.430-445. |
| 3066A | A1 | Mintrop, H (2003) The Old and New Face of Civic Education: expert, teacher, and student views. <i>European Educational Research Journal</i> . Vol.2, No.3, pp.446-454. |
| 3067 | M | Blair, H (2003) Jump-Starting Democracy: Adult Civic Education and Democratic Participation in Three Countries. <i>Democratization</i> . Vol.10, No.1, pp.53-76. |
| 3068 | N | Levstik, L S (2002) Introduction. <i>Elementary School Journal</i> . Vol.103, No.2, p.93. |
| 3069A | A1 | Torney-Purta, J (2002) The School's Role in Developing Civic Engagement: A Study of Adolescents in Twenty-Eight Countries. <i>Applied Developmental Science</i> . Vol.6, No.4, pp.203-212. |
| 3070A | A2 | Highlights of U.S. Results from the IEA Civic Education Study. <i>Social Studies</i> . Vol.92, No.6, pp.274-275. |
| 3071 | M | Grossman, D & Cheung, K C (2001) What kind of citizens are required for the 21st century: A case study of pre-service civic education teachers in Hong Kong and China. <i>Pacific-Asian Education Journal</i> . Vol.13, No.1, pp.20-37. |
| 3072A | A1 | Osler, A & Starkey, H (2001) Citizenship Education and National Identities in France and England: inclusive or exclusive? <i>Oxford Review of Education</i> . Vol.27, No.2, pp.287-305. |
| 3073 | M | Ulrich, W (2001) The quest for competence in systemic research and practice. <i>Systems Research & Behavioral Science</i> . Vol.18, No.1, pp.3-28. |
| 3074B | B | Torney-Purta, J (2000) An International Perspective on the NAEP Civics Report Card. <i>Social Studies</i> . Vol.91, No.4, pp.148-151. |
| 3075 | M | Wray, H (2000) The fall of moral education and the rise and decline of civics education and social studies in occupied and independent Japan. <i>Japan Forum</i> . Vol.12, No.1, pp.15-41. |
| 3076 | M | Sansone, S C (1999) Get your students involved in civics. <i>Social Education</i> . Vol.63, No.4, pp.228-233. |
| 3077 | M | Hahn, C L (1999) Citizenship Education: an empirical study of policy, practices and outcomes. <i>Oxford Review of Education</i> . Vol.25, No.1/2, pp.231-250. |
| 3078A | NA | Hahn, C L & Torney-Purta, J (1999) The IEA Civic Education Project: National and International Perspectives. <i>Social Education</i> . Vol.63, No.7, pp.425-531. |
| 3079 | M | Kwan-Choi, T T (1998) Preparing Students for Citizenship? Political Education in Hong Kong. <i>Education & Society</i> . Vol.16, No.2, pp.5-15. |
| 3080 | M | Parker-Gwin, R & Mabry, J B (1998) Service Learning as Pedagogy and Civic Education: Comparing Outcomes for Three Models. <i>Teaching Sociology</i> . Vol.26, No.4, pp.276-291. |
| 3081 | M | Cheung, C-K & Leung, M (1998) From civic education to general studies: The implementation of political education into the... <i>Compare: A Journal of Comparative Education</i> . Vol.28, No.1, pp.47-57. |
| 3082 | M | Vinson, K D (1998) The Problematics of Character Education and Civic Virtue: A Critical Response to the NCSS Position Statement. <i>Social Education</i> . Vol.62, No.2, pp.112-115. |
| 4001 | M | Myers, J P & Zaman, H A (2009) Negotiating the Global and National: Immigrant and Dominant-Culture Adolescents' Vocabularies of Citizenship in a Transnational World. <i>Teachers College Record</i> . Vol.111, No.11, pp.2589-2625. |
| 4002 | M | Law, W-W & Ng, H M (2009) Globalization and Multileveled Citizenship Education: A Tale of Two Chinese Cities, Hong Kong and Shanghai. <i>Teachers College Record</i> . Vol.111, No.4, pp.851-892. |
| 4003 | M | Quaynor, L (2012) Citizenship Education in Post-Conflict Contexts: A Review of the Literature. <i>Education, Citizenship and Social Justice</i> . Vol.7, No.1, pp.33-57. |
| 4004A | A2 | Wainaina, P K, Amot, M & Chege, F (2011) Developing Ethical and Democratic Citizens in a Post-Colonial Context: Citizenship Education in Kenya. <i>Educational Research</i> . Vol.53, No.2, pp.179-192. |
| 4005 | M | Alviar-Martin, T (2011) Reconciling Multiple Conceptions of Citizenship: International School Teachers' Beliefs and Practice. <i>Journal of Education</i> . Vol.191, No.3, pp.39-49. |
| 4006A | A1 | Kennedy, K J, Kuang, X & Chow, J K F (2013) Exploring Asian Students' Citizenship Values and Their Relationship to Civic Knowledge and School Participation. <i>Educational Psychology</i> . Vol.33, No.3, pp.233-254. |
| 4007A | A2 | Myers, J P (2010) 'To benefit the world by whatever means possible': adolescents' constructed meanings for global citizenship. <i>British Educational Research Journal</i> . Vol.36, No.3, pp.483-502. |
| 4008A | A1 | Verhaegen, S, Hooghe, M & Meeusen, C (2013) Opportunities to learn about Europe at school. A |

| # | C | Reference |
|-------|----|---|
| | | comparative analysis among European adolescents in 21 European member states. <i>Journal of Curriculum Studies</i> . Vol.45, No.6, pp.838-864. |
| 4009A | A2 | Gilleece, L & Cosgrove, J (2012) Student Civic Participation in School: What Makes a Difference in Ireland? <i>Education, Citizenship and Social Justice</i> . Vol.7, No.3, pp.225-239. |
| 4010A | NA | Ajiboye, J O (2009) Children Media Usage, Civic Interest and Political-Related Practices among Botswana Primary School Pupils: An Experience in Informal Civic Education. <i>Journal of Social Sciences</i> . Vol.20, No.3, pp.189-196. |
| 4011A | A2 | Walter, F, Rosenberger, S & Ptaszynska, A (2013) Challenging the boundaries of democratic inclusion? Young people's attitudes about the distribution of voting rights. <i>Citizenship Studies</i> . Vol.17, No.3/4, pp.464-478. |
| 4012 | M | Tarrant, M A (2010) A Conceptual Framework for Exploring the Role of Studies Abroad in Nurturing Global Citizenship. <i>Journal of Studies in International Education</i> . Vol.14, No.5, pp.433-451. |
| 4013 | M | Berkson, S, Espinola, J, Corso, K, Cabral, H, McGowan, R & Chomitz, V (2013) Reliability of Height and Weight Measurements Collected by Physical Education Teachers for a School-Based Body Mass Index Surveillance and Screening System. <i>Journal of School Health</i> . Vol.83, No.1, pp.21-27. |
| 4014 | M | Shaklee, B & Merz, S (2012) Intercultural communication competency for international educators. <i>International Schools Journal</i> . Vol.32, No.1, pp.13-20. |
| 4015 | M | Design and development of rubrics to improve assessment outcomes: A pilot study in a Master's level business program in India. (2011) <i>Quality Assurance in Education: An International Perspective</i> . Vol.19, No.1, pp.84-104. |
| 4016 | M | Laopongharn, W & Sercombe, P (2009) What Relevance does Intercultural Communication have to Language Education in Thailand? <i>Annual Review of Education, Communication & Language Sciences</i> . Vol.6, pp.59-83. |

APPENDIX II: MAJOR PUBLICATIONS FROM ILSA

IEA Publications

| | |
|---|---|
| Pilot Twelve-Country Study | Foshay, A.W., Thorndike, R.L., Hotyat, F., Pidgeon, D.A., & Walker, D.A. (1962). Educational achievements of thirteen-year-olds in twelve countries: Results of an international research project, 1959–1961. Hamburg: UNESCO Institute for Education. |
| First International Mathematics Study, FIMS | Husén, T. (Ed.). (1967). International study of achievement in mathematics: A comparison of twelve countries (Vols. 1–2). Stockholm: Almqvist & Wiksell. Postlethwaite, N. (1967). School organization and student achievement: A study based on achievement in mathematics in twelve countries. Stockholm: Almqvist & Wiksell |
| The Six-Subject Study: Science First International Science Study, FISS | Bloom, B.S. (1969). Cross-national study of educational attainment: Stage I of the I.E.A. investigation in six subject areas (Vols. 1–2). Washington, DC: Office of Education (DHEW). Comber, L.C., & Keeves, J.P. (1973). Science education in nineteen countries: An empirical study. Stockholm: Almqvist & Wiksell. Passow, A.H., Noah, H.J., Eckstein, M.A., & Mallea, J.R. (1976). The national case study: An empirical comparative study of twenty-one educational systems. Stockholm: Almqvist & Wiksell. Peaker, G.F. (1975). An empirical study of education in twenty-one countries: A technical report. Stockholm: Almqvist & Wiksell. Walker, D.A. (1976). The IEA Six Subject Survey: An empirical study of education in twenty-one countries. Stockholm: Almqvist & Wiksell |
| The Six-Subject Study: Reading Comprehension | Bloom, B.S. (1969). Cross-national study of educational attainment: Stage I of the I.E.A. investigation in six subject areas (Vols. 1–2). Washington, DC: Office of Education (DHEW). Passow, A.H., Noah, H.J., Eckstein, M.A., & Mallea, J.R. (1976). The national case study: An empirical comparative study of twenty-one educational systems. Stockholm: Almqvist & Wiksell. Peaker, G.F. (1975). An empirical study of education in twenty-one countries: A technical report. Stockholm: Almqvist & Wiksell. Thorndike, R.L. (1973). Reading comprehension education in fifteen countries: An empirical study. Stockholm: Almqvist & Wiksell. Walker, D.A. (1976). The IEA Six Subject Survey: An empirical study of education in twenty-one countries. Stockholm: Almqvist & Wiksell. |
| The Six-Subject Study: Literature Education | Bloom, B.S. (1969). Cross-national study of educational attainment: Stage I of the I.E.A. investigation in six subject areas (Vols. 1–2). Washington, DC: Office of Education (DHEW). Passow, A.H., Noah, H.J., Eckstein, M.A., & Mallea, J.R. (1976). The national case study: An empirical comparative study of twenty-one educational systems. Stockholm: Almqvist & Wiksell. Peaker, G.F. (1975). An empirical study of education in twenty-one countries: A technical report. Stockholm: Almqvist & Wiksell. Purves, A.C. (1973). Literature education in ten countries: An empirical study. Stockholm: Almqvist & Wiksell. Walker, D.A. (1976). The IEA Six Subject Survey: An empirical study of education in twenty-one countries. Stockholm: Almqvist & Wiksell. |
| The Six-Subject Study: English as a Foreign Language | Bloom, B.S. (1969). Cross-national study of educational attainment: Stage I of the I.E.A. investigation in six subject areas (Vols. 1–2). Washington, DC: Office of Education (DHEW). Lewis, E.G., & Massad, C.E. (1975). The teaching of English as a foreign language in ten countries. Stockholm: Almqvist & Wiksell. Passow, A.H., Noah, H.J., Eckstein, M.A., & Mallea, J.R. (1976). The national case study: An empirical comparative study of twenty-one educational systems. Stockholm: Almqvist & Wiksell. Peaker, G.F. (1975). An empirical study of education in twenty-one countries: A technical report. Stockholm: Almqvist & Wiksell. Walker, D.A. (1976). The IEA Six Subject Survey: An empirical study of education in twenty-one countries. Stockholm: Almqvist & Wiksell. |
| The Six-Subject Study: French as a Foreign Language | Bloom, B.S. (1969). Cross-national study of educational attainment: Stage I of the I.E.A. investigation in six subject areas (Vols. 1–2). Washington, DC: Office of Education (DHEW). Carroll, J.B. (1975). The teaching of French as a foreign language in eight countries. Stockholm: Almqvist & Wiksell. Passow, A.H., Noah, H.J., Eckstein, M.A., & Mallea, J.R. (1976). The national case study: An empirical comparative study of twenty-one educational systems. Stockholm: Almqvist & Wiksell. Peaker, G.F. (1975). An empirical study of education in twenty-one countries: A technical report. Stockholm: Almqvist & Wiksell. Walker, D.A. (1976). The IEA Six Subject Survey: An empirical study of education in twenty-one countries. Stockholm: Almqvist & Wiksell. |
| The Six-Subject Study: Civic Education | Bloom, B.S. (1969). Cross-national study of educational attainment: Stage I of the I.E.A. investigation in six subject areas (Vols. 1–2). Washington, DC: Office of Education (DHEW). |

| | |
|--|---|
| | <p>Oppenheim, A.N., & Torney, J. (1974). <i>The measurement of children's civic attitudes in different nations</i>. Stockholm: Almqvist & Wiksell.</p> <p>Passow, A.H., Noah, H.J., Eckstein, M.A., & Mallea, J.R. (1976). <i>The national case study: An empirical comparative study of twenty-one educational systems</i>. Stockholm: Almqvist & Wiksell.</p> <p>Peaker, G.F. (1975). <i>An empirical study of education in twenty-one countries: A technical report</i>. Stockholm: Almqvist & Wiksell.</p> <p>Torney, J.V., Oppenheim, A.N., & Farnen, R.F. (1975). <i>Civic education in ten countries: An empirical study</i>. Stockholm: Almqvist & Wiksell.</p> <p>Walker, D.A. (1976). <i>The IEA Six Subject Survey: An empirical study of education in twenty-one countries</i>. Stockholm: Almqvist & Wiksell.</p> |
| Second International Mathematics Study, SIMS | <p>Burstein, L. (Ed.). (1993). <i>The IEA Study of Mathematics III: Student growth and classroom processes</i>. Oxford: Pergamon Press.</p> <p>Garden, R.A. (1987). <i>Second IEA Mathematics Study: Sampling report</i>. Washington, DC: Center for Education Statistics.</p> <p>Jaji, G. (1986). <i>Second International Mathematics Study: The use of calculators and computers in mathematics classes in twenty countries—a source document</i>. Washington, DC: Center for Education Statistics.</p> <p>Livingstone, I.D. (1986). <i>Second International Mathematics Study: Perceptions of the intended and implemented mathematics curriculum</i>. Washington, DC: Center for Education Statistics.</p> <p>Pelgrum, W.J., Eggen, T., & Plomp, T. (1986). <i>Second International Mathematics Study: The implemented and attained mathematics curriculum—a comparison of eighteen countries</i>. Washington, DC: Center for Education Statistics.</p> <p>Robitaille, D.F., & Garden, R.A. (Eds.). (1989). <i>The IEA Study of Mathematics II: Contexts and outcomes of school mathematics</i>. Oxford: Pergamon Press.</p> <p>Travers, K.J., & Westbury, I. (1989). <i>The IEA Study of Mathematics I: Analysis of mathematics curricula</i>. Oxford: Pergamon Press.</p> <p>Travers, K.J., & Westbury, I. (1989). <i>Supplement to the IEA Study of Mathematics I: Analysis of mathematics curricula</i>. Urbana, IL: University of Illinois.</p> <p>Westbury, I., & Travers, K. (1990). <i>Second International Mathematics Study: Studies</i>. Urbana, IL: University of Illinois.</p> <p>Wolfe, R.G. (1987). <i>Second International Mathematics Study: Training manual for use of the databank of the longitudinal, classroom process surveys for Population A in the IEA Second International Mathematics Study</i>. Washington, DC: Center for Education Statistics.</p> |
| Classroom Environment Study | <p>Anderson, L.W., Ryan, D.W., & Shapiro, B.J. (Eds.). (1989). <i>The IEA Classroom Environment Study</i>. Oxford: Pergamon Press.</p> |
| Second International Science Study, SISS | <p>IEA. (1988). <i>Science achievement in seventeen countries: A preliminary report</i>. Oxford: Pergamon Press.</p> <p>Keeves, J.P. (1992). <i>Learning science in a changing world: Cross-national studies of science achievement, 1970 to 1984</i>. The Hague: IEA.</p> <p>Keeves, J.P. (Ed.). (1992). <i>The IEA Study of Science III: Changes in science education and achievement, 1970 to 1984</i>. Oxford: Pergamon Press.</p> <p>Postlethwaite, T.N., & Wiley, D.E. (Eds.). (1992). <i>The IEA Study of Science II: Science achievement in twenty-three countries</i>. Oxford: Pergamon Press.</p> <p>Rosier, M.J., & Keeves, J.P. (Eds.). (1991). <i>The IEA Study of Science I: Science education and curricula in twenty-three countries</i>. Oxford: Pergamon Press.</p> |
| Written Composition Study | <p>Gorman, T.P., Purves, A.C., & Degenhart, R.E. (Eds.). (1988). <i>The IEA Study of Written Composition I: The international writing tasks and scoring scales</i>. Oxford: Pergamon Press.</p> <p>Purves, A.C. (Ed.). (1992). <i>The IEA Study of Written Composition II: Education and performance in fourteen countries</i>. Oxford: Pergamon Press.</p> <p>Saari, H. (1991). <i>Writing curricula in sixteen countries: International study in written composition—IEA</i>. Jyväskylä, Finland: University of Jyväskylä.</p> |
| Reading Literacy Study | <p>Binkley, M., Rust, K., & Williams, T. (Eds.). (1996). <i>Reading literacy in an international perspective: Collected papers from the IEA Reading Literacy Study</i>. Washington, DC: NCES.</p> <p>Binkley, M., Rust, K., & Winglee, M. (Eds.). (1995). <i>Methodological issues in comparative educational studies: The case of the IEA Reading Literacy Study</i>. Washington, DC: NCES.</p> <p>Elley, W.B. (1992). <i>How in the world do students read? IEA Study of Reading Literacy</i>. The Hague: IEA.</p> <p>Elley, W.B. (Ed.). (1994). <i>The IEA Study of Reading Literacy: Achievement and instruction in thirty-two school systems</i>. Oxford: Pergamon Press.</p> <p>Lundberg, I., & Linnakylä, P. (1993). <i>Teaching reading around the world: IEA Study of Reading Literacy</i>. The Hague: IEA.</p> <p>Papanastasiou, C., & Froese, V. (2002). <i>Reading literacy in 14 countries</i>. Lefkosia, Cyprus: University of Cyprus.</p> <p>Postlethwaite, T.N., & Ross, K.N. (1992). <i>Effective schools in reading: Implications for</i></p> |

| | |
|---|---|
| | <p>educational planners. An exploratory study. The Hague: IEA.</p> <p>Wagemaker, H. (Ed.). (1996). Are girls better readers? Gender differences in reading literacy in 32 countries. Amsterdam: IEA.</p> <p>Wolf, R.M. (Ed.). (1995). The IEA Reading Literacy Study: Technical report. The Hague: IEA.</p> |
| Computer in Education/Information Technology, COMPED | <p>Pelgrum, W.J., & Plomp, T. (1991). The use of computers in education worldwide: Results from the IEA 'Computers in Education' survey in 19 educational systems. Oxford: Pergamon Press.</p> <p>Pelgrum, W.J., & Plomp, T. (Eds.). (1993). The IEA Study of Computers in Education: Implementation of an innovation in 21 education systems. Oxford: Pergamon Press.</p> <p>Pelgrum, W.J., Reinen, I.A.M.J., & Plomp, T. (Eds.). (1993). Schools, teachers, students and computers: A cross-national perspective. IEA-Comped Study Stage 2. Enschede, Netherlands: University of Twente.</p> <p>Plomp, T., Anderson, R.E., & Kontogiannopoulou-Polydorides, G. (Eds.). (1996). Cross national policies and practices on computers in education. Dordrecht: Kluwer Academic Publishers</p> |
| Language Education Study, LES | <p>Dickson, P., & Cumming, A. (Eds.). (1996). Profiles of language education in 25 countries: Overview of Phase 1 of the IEA Language Education Study. Slough, UK: National Foundation for Educational Research.</p> |
| Third International Mathematics and Science Study, TIMSS 1995 | <p>Beaton, A.E., Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., Smith, T.A., & Kelly, D.L. (1996). Science achievement in the middle school years: IEA's Third International Mathematics and Science Study. Chestnut Hill, MA: Boston College.</p> <p>Beaton, A.E., Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., Kelly, D.L., & Smith, T.A. (1996). Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study. Chestnut Hill, MA: Boston College.</p> <p>Gonzalez, E.J., & Smith, T.A. (Eds.). (1997). User guide for the TIMSS international database: Primary and middle school years. Chestnut Hill, MA: Boston College.</p> <p>Gonzalez, E.J., Smith, T.A., & Sibberns, H. (Eds.). (1998). User guide for the TIMSS international database: Final year of secondary school. Chestnut Hill, MA: Boston College.</p> <p>Harmon, M., Smith, T.A., Martin, M.O., Kelly, D.L., Beaton, A.E., Mullis, I.V.S., Gonzalez, E.J., & Orpwood, G. (1997). Performance assessment in IEA's Third International Mathematics and Science Study. Chestnut Hill, MA: Boston College.</p> <p>Howson, G. (1995). Mathematics textbooks: A comparative study of Grade 8 texts. TIMSS monograph no. 3. Vancouver, Canada: Pacific Educational Press.</p> <p>Martin, M.O., & Kelly, D.L. (Eds.). (1996). Third International Mathematics and Science Study technical report, Volume I: Design and development. Chestnut Hill, MA: Boston College.</p> <p>Martin, M.O., & Kelly, D.L. (Eds.). (1997). Third International Mathematics and Science Study technical report, Volume II: Implementation and analysis—primary and middle school years. Chestnut Hill, MA: Boston College.</p> <p>Martin, M.O., & Kelly, D.L. (Eds.). (1998). Third International Mathematics and Science Study technical report, Volume III: Implementation and analysis—final year of secondary school. Chestnut Hill, MA: Boston College.</p> <p>Martin, M.O., & Mullis, I.V.S. (Eds.). (1996). Third International Mathematics and Science Study: Quality assurance in data collection. Chestnut Hill, MA: Boston College.</p> <p>Martin, M.O., Mullis, I.V.S., Beaton, A.E., Gonzalez, E.J., Smith, T.A., & Kelly, D.L. (1997). Science achievement in the primary school years: IEA's Third International Mathematics and Science Study. Chestnut Hill, MA: Boston College.</p> <p>Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., Smith, T.A., & Kelly, D.L. (1999). School contexts for learning and instruction: IEA's Third International Mathematics and Science Study. Chestnut Hill, MA: Boston College.</p> <p>Martin, M.O., Mullis, I.V.S., Gregory, K.D., Hoyle, C., & Shen, C. (2000). Effective schools in science and mathematics: IEA's Third International Mathematics and Science Study. Chestnut Hill, MA: Boston College.</p> <p>Mullis, I.V.S., Martin, M.O., Beaton, A.E., Gonzalez, E.J., Kelly, D.L., & Smith, T.A. (1997). Mathematics achievement in the primary school years: IEA's Third International Mathematics and Science Study. Chestnut Hill, MA: Boston College.</p> <p>Mullis, I.V.S., Martin, M.O., Beaton, A.E., Gonzalez, E.J., Kelly, D.L., & Smith, T.A. (1998). Mathematics and science achievement in the final year of secondary school: IEA's Third International Mathematics and Science Study. Chestnut Hill, MA: Boston College.</p> <p>Mullis, I.V.S., Martin, M.O., Fierros, E.G., Goldberg, A.L., & Stemler, S.E. (2000). Gender differences in achievement: IEA's Third International Mathematics and Science Study. Chestnut Hill, MA: Boston College.</p> <p>Orpwood, G., & Garden, R.A. (1998). Assessing mathematics and science literacy: TIMSS monograph no. 4. Vancouver, Canada: Pacific Educational Press.</p> <p>Robitaille, D.F. (Ed.). (1997). National contexts for mathematics and science education: An encyclopedia of the education systems participating in TIMSS. Vancouver, Canada: Pacific</p> |

| | |
|---|--|
| | <p>Educational Press.</p> <p>Robitaille, D.F., & Beaton, A.E. (Eds.). (2002). Secondary analysis of the TIMSS data. Dordrecht: Kluwer Academic Publishers.</p> <p>Robitaille, D.F., Beaton, A.E., & Plomp, T. (Eds.). (2000). The impact of TIMSS on the teaching & learning of mathematics & science. Vancouver, Canada: Pacific Educational Press.</p> <p>Robitaille, D.F., & Garden, R.A. (Eds.). (1996). Research questions & study design: TIMSS monograph no. 2. Vancouver, Canada: Pacific Educational Press.</p> <p>Robitaille, D.F., Schmidt, W.H., Raizen, S., McKnight, C., Britton, E., & Nicol, C. (1993). Curriculum frameworks for mathematics and science: TIMSS monograph no. 1. Vancouver, Canada: Pacific Educational Press.</p> <p>Schmidt, W.H., McKnight, C.C., Valverde, G.A., Houang, R.T., & Wiley, D.E. (1997). Many visions, many aims, Volume 1: A cross-national investigation of curricular intentions in school mathematics. Dordrecht: Kluwer Academic Publishers.</p> <p>Schmidt, W.H., Raizen, S.A., Britton, E.D., Bianchi, L.J., & Wolfe, R.G. (1997). Many visions, many aims, Volume 2: A cross-national investigation of curricular intentions in school science. Dordrecht: Kluwer Academic Publishers</p> |
| Second Information Technology in Education Study Module 1, SITES-M1 | <p>Pelgrum, W.J., & Anderson, R.E. (Eds.). (2001). ICT and the emerging paradigm for life-long learning: An IEA educational assessment of infrastructure, goals, and practices in twenty-six countries (2nd ed.). Amsterdam: IEA</p> |
| Third International Mathematics and Study Repeat, TIMSS-R 1999 | <p>Gonzalez, E.J., & Miles, J.A. (Eds.). (2001). TIMSS 1999 user guide for the international database. Chestnut Hill, MA: Boston College.</p> <p>Martin, M.O., Gregory, K.D., & Stemler, S.E. (Eds.). (2000). TIMSS 1999 technical report. Chestnut Hill, MA: Boston College.</p> <p>Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., Gregory, K.D., Smith, T.A., Chrostowski, S.J., Garden, R.A., & O'Connor, K.M. (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.</p> <p>Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., Gregory, K.D., Garden, R.A., O'Connor, K.M., Chrostowski, S.J., & Smith, T.A. (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.</p> |
| TIMSS 1999 Video Study | <p>Garnier, H.E., Lemmens, M., Druker, S.L., & Roth, K.J. (2011). Third International Mathematics and Science Study 1999 Video Study technical report, Volume 2: Science. Washington, DC: NCES.</p> <p>Hiebert, J., Gallimore, R., Garnier, H., Givvin, K.B., Hollingsworth, H., Jacobs, J., Chui, A.M.-Y., Wearne, D., Smith, M., Kersting, N., Manaster, A., Tseng, E., Etterbeek, W., Manaster, C., Gonzales, P., & Stigler, J. (2003). Teaching mathematics in seven countries: Results from the TIMSS 1999 Video Study. Washington, DC: NCES.</p> <p>Jacobs, J., Garnier, H., Gallimore, R., Hollingsworth, H., Givvin, K.B., Rust, K., Kawanaka, T., Smith, M., Wearne, D., Manaster, A., Etterbeek, W., Hiebert, J., & Stigler, J. (2003). Third International Mathematics and Science Study 1999 Video Study technical report, Volume 1: Mathematics. Washington, DC: NCES.</p> <p>Roth, K.J., Druker, S.L., Garnier, H.E., Lemmens, M., Chen, C., Kawanaka, T., Rasmussen, D., Trubacova, S., Warvi, D., Okamoto, Y., Gonzales, P., Stigler, J., & Gallimore, R. (2006). Teaching science in five countries: Results from the TIMSS 1999 Video Study. Washington, DC: NCES.</p> |
| Civic Education Study, CIVED | <p>Amadeo, J.-A., Torney-Purta, J., Lehmann, R., Husfeldt, V., & Nikolova, R. (2002). Civic knowledge and engagement: An IEA study of upper secondary students in sixteen countries. Amsterdam: IEA.</p> <p>Schulz, W., & Sibberns, H. (Eds.). (2004). IEA Civic Education Study technical report. Amsterdam: IEA.</p> <p>Sibberns, H. (2005). IEA Civic Education Study user guide for the international database. Amsterdam: IEA.</p> <p>Steiner-Khamsi, G., Torney-Purta, J., & Schwille, J. (Eds.). (2002). New paradigms and recurring paradoxes in education for citizenship: An international comparison. Oxford: Elsevier Science.</p> <p>Torney-Purta, J., Lehmann, R., Oswald, H., & Schulz, W. (2001). Citizenship and education in twenty-eight countries: Civic knowledge and engagement at age fourteen. Amsterdam: IEA.</p> <p>Torney-Purta, J., Schwille, J., & Amadeo, J.-A. (Eds.). (1999). Civic education across countries: Twenty-four national case studies from the IEA Civic Education Project. Amsterdam: IEA.</p> |
| Second Information Technology in Education Study Module 2, SITES-M2 | <p>Kozma, R.B. (Ed.). (2003). Technology, innovation, and educational change: A global perspective. A report of the Second Information Technology in Education Study Module 2. Eugene, OR: ISTE.</p> <p>Plomp, T., Anderson, R.E., Law, N., & Quale, A. (Eds.). (2003). Cross-national information and communication technology policies and practices in education. Greenwich, CT: Information</p> |

| | |
|--|--|
| | Age Publishing. |
| Progress in International Reading Literacy Study, PIRLS 2001 | <p>Campbell, J.R., Kelly, D.L., Mullis, I.V.S., Martin, M.O., & Sainsbury, M. (2001). Framework and specifications for PIRLS assessment 2001 (2nd ed.). Chestnut Hill, MA: Boston College.</p> <p>Gonzalez, E.J., & Kennedy, A.M. (Eds.). (2003). PIRLS 2001 user guide for the international database. Chestnut Hill, MA: Boston College.</p> <p>Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., & Kennedy, A.M. (2003). Trends in children's reading literacy achievement 1991–2001: IEA's repeat in nine countries of the 1991 Reading Literacy Study. Chestnut Hill, MA: Boston College.</p> <p>Martin, M.O., Mullis, I.V.S., & Kennedy, A.M. (Eds.). (2003). PIRLS 2001 technical report. Chestnut Hill, MA: Boston College.</p> <p>Mullis, I.V.S., Martin, M.O., & Gonzalez, E.J. (2004). International achievement in the processes of reading comprehension: Results from PIRLS 2001 in 35 countries. Chestnut Hill, MA: Boston College.</p> <p>Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., & Kennedy, A.M. (2003). PIRLS 2001 international report: IEA's study of reading literacy achievement in primary schools in 35 countries. Chestnut Hill, MA: Boston College.</p> <p>Mullis, I.V.S., Martin, M.O., Kennedy, A.M., & Flaherty, C.L. (Eds.). (2002). PIRLS 2001 encyclopedia: A reference guide to reading education in the countries participating in IEA's Progress in International Reading Literacy Study (PIRLS). Chestnut Hill, MA: Boston College.</p> <p>Schwippert, K. (Ed.). (2007). Progress in reading literacy: The impact of PIRLS 2001 in 13 countries. Münster: Waxmann.</p> |
| Pre-Primary Education Project, PPP | <p>High/Scope Foundation. (1994). Sights and sounds of children: High/Scope international videotape series [Set of 15 videotapes]. Ypsilanti, MI: High/Scope Press.</p> <p>Montie, J.E., Xiang, Z., & Schweinhart, L.J. (Eds.). (2007). The role of preschool experience in children's development: Longitudinal findings from 10 countries. The IEA Preprimary Project Phase 3. Ypsilanti, MI: High/Scope Press.</p> <p>Olmsted, P.P., & Montie, J. (Eds.). (2001). Early childhood settings in 15 countries: What are their structural characteristics? Ypsilanti, MI: High/Scope Press.</p> <p>Olmsted, P.P., & Weikart, D.P. (Eds.). (1989). How nations serve young children: Profiles of child care and education in 14 countries. Ypsilanti, MI: High/Scope Press.</p> <p>Olmsted, P.P., & Weikart, D.P. (Eds.) (1994). Families speak: Early childhood care and education in 11 countries. The IEA Preprimary Project Phase 1 [Softcover edition]. Ypsilanti, MI: High/Scope Press.</p> <p>Olmsted, P.P., & Weikart, D.P. (Eds.) (1995). The IEA Preprimary Study: Early childhood care and education in 11 countries [Hardcover edition]. Oxford: Pergamon Press.</p> <p>Weikart, D.P. (Ed.). (1999). What should young children learn? Teacher and parent views in 15 countries. The IEA Preprimary Project Phase 2. Ypsilanti, MI: High/Scope Press.</p> <p>Weikart, D.P., Olmsted, P.P., & Montie, J. (Eds.). (2003). A world of preschool experience: Observations in 15 countries. The IEA Preprimary Project Phase 2. Ypsilanti, MI: High/Scope Press.</p> |
| Trends in Mathematics and Science Study, TIMSS 2003 | <p>Martin, M.O. (Ed.). (2005). TIMSS 2003 user guide for the international database. Chestnut Hill, MA: Boston College.</p> <p>Martin, M.O., Mullis, I.V.S., & Chrostowski, S.J. (Eds.). (2004). TIMSS 2003 technical report. Chestnut Hill, MA: Boston College.</p> <p>Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., & Chrostowski, S.J. (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.</p> <p>Mullis, I.V.S., Martin, M.O., & Foy, P. (2005). IEA's TIMSS 2003 international report on achievement in the mathematics cognitive domains: Findings from a developmental project. Chestnut Hill, MA: Boston College.</p> <p>Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., & Chrostowski, S.J. (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.</p> <p>Mullis, I.V.S., Martin, M.O., Smith, T.A., Garden, R.A., Gregory, K.D., Gonzalez, E.J., Chrostowski, S.J., & O'Connor, K.M. (2002). Marcos teóricos y especificaciones de evaluación de TIMSS 2003 [TIMSS assessment frameworks and specifications 2003]. Madrid: INCE.</p> <p>Mullis, I.V.S., Martin, M.O., Smith, T.A., Garden, R.A., Gregory, K.D., Gonzalez, E.J., Chrostowski, S.J., & O'Connor, K.M. (2003). TIMSS assessment frameworks and specifications 2003 (2nd ed.). Chestnut Hill, MA: Boston College.</p> |
| Second Information Technology in Education Study, SITES 2006 | <p>Brese, F., & Carstens, R. (Eds.). (2009). Second Information Technology in Education Study: SITES 2006 user guide for the international database. Amsterdam: IEA.</p> <p>Carstens, R., & Pelgrum, W.J. (Eds.). (2009). Second Information Technology in Education Study: SITES 2006 technical report. Amsterdam: IEA.</p> <p>Law, N., Pelgrum, W.J., & Plomp, T. (Eds.). (2008). Pedagogy and ICT use in schools around the world: Findings from the IEA SITES 2006 study. Hong Kong: CERCSpringer.</p> |

| | |
|---|---|
| | <p>Plomp, T., Anderson, R.E., Law, N., & Quale, A. (Eds.). (2009). <i>Cross-national information and communication technology: Policies and practices in education</i> (Rev. 2nd ed.). Charlotte, NC: Information Age Publishing.</p> |
| Progress in International Reading Literacy Study, PIRLS 2006 | <p>Foy, P., & Kennedy, A.M. (Eds.). (2008). <i>PIRLS 2006 user guide for the international database</i>. Chestnut Hill, MA: Boston College.</p> <p>Kennedy, A.M., Mullis, I.V.S., Martin, M.O., & Trong, K.L. (Eds.). (2007). <i>PIRLS 2006 encyclopedia: A guide to reading education in the forty PIRLS 2006 countries</i>. Chestnut Hill, MA: Boston College.</p> <p>Martin, M.O., Mullis, I.V.S., & Kennedy, A.M. (Eds.). (2007). <i>PIRLS 2006 technical report</i>. Chestnut Hill, MA: Boston College.</p> <p>Mullis, I.V.S., Kennedy, A.M., Martin, M.O., & Sainsbury, M. (2006). <i>PIRLS 2006 assessment framework and specifications: Progress in International Reading Literacy Study</i> (2nd ed.). Chestnut Hill, MA: Boston College.</p> <p>Mullis, I.V.S., Kennedy, A.M., Martin, M.O., & Sainsbury, M. (2006). <i>PIRLS 2006 Marcos teóricos y especificaciones de evaluación: Estudio Internacional de Progreso en Comprensión Lectora [PIRLS 2006 assessment framework and specifications: Progress in International Reading Literacy Study]</i> (2nd ed.). INECSE.</p> <p>Mullis, I.V.S., Martin, M.O., Kennedy, A.M., & Foy, P. (2007). <i>PIRLS 2006 international report: IEA's Progress in International Reading Literacy Study in primary schools in 40 countries</i>. Chestnut Hill, MA: Boston College.</p> <p>Schwippert, K., & Lenkeit, J. (Eds.). (2012). <i>Progress in reading literacy in national and international context: The impact of PIRLS 2006 in 12 countries</i>. Münster: Waxmann.</p> |
| Trends in Mathematics and Science Study, TIMSS 2007 | <p>Foy, P., & Olson, J.F. (Eds.). (2009). <i>TIMSS 2007 user guide for the international database</i>. Chestnut Hill, MA: Boston College.</p> <p>Martin, M.O., Mullis, I.V.S., & Foy, P. (2008). <i>TIMSS 2007 international science report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades</i>. Chestnut Hill, MA: Boston College.</p> <p>Mullis, I.V.S., Martin, M.O., & Foy, P. (2008). <i>TIMSS 2007 international mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades</i>. Chestnut Hill, MA: Boston College.</p> <p>Mullis, I.V.S., Martin, M.O., Olson, J.F., Berger, D.R., Milne, D., & Stanco, G.M. (Eds.) (2008). <i>TIMSS 2007 encyclopedia: A guide to mathematics and science education around the world</i> (Vols. 1–2). Chestnut Hill, MA: Boston College.</p> <p>Mullis, I.V.S., Martin, M.O., Ruddock, G.J., O'Sullivan, C.Y., Arora, A., & Erberber, E. (2005). <i>TIMSS 2007 assessment frameworks</i>. Chestnut Hill, MA: Boston College.</p> <p>Olson, J.F., Martin, M.O., & Mullis, I.V.S. (Eds.). (2008). <i>TIMSS 2007 technical report</i>. Chestnut Hill, MA: Boston College.</p> |
| TIMSS Advanced 2008 | <p>Arora, A., Foy, P., Martin, M.O., & Mullis, I.V.S. (Eds.). (2009). <i>TIMSS Advanced 2008 technical report</i>. Chestnut Hill, MA: Boston College.</p> <p>Foy, P., & Arora, A. (Eds.). (2009). <i>TIMSS Advanced 2008 user guide for the international database</i>. Chestnut Hill, MA: Boston College.</p> <p>Garden, R.A., Lie, S., Robitaille, D.F., Angell, C., Martin, M.O., Mullis, I.V.S., Foy, P., & Arora, A. (2006). <i>TIMSS Advanced 2008 assessment frameworks</i>. Chestnut Hill, MA: Boston College.</p> <p>Garden, R.A., Lie, S., Robitaille, D.F., Angell, C., Martin, M.O., Mullis, I.V.S., Foy, P., & Arora, A. (2010). <i>TIMSS Advanced 2008: Quadri di riferimento per la valutazione [TIMSS Advanced 2008 assessment frameworks]</i>. Frascati, Italy: INVALSI.</p> <p>Mullis, I.V.S., Martin, M.O., Robitaille, D.F., & Foy, P. (2009). <i>TIMSS Advanced 2008 international report: Findings from IEA's study of achievement in advanced mathematics and physics in the final year of secondary school</i>. Chestnut Hill, MA: Boston College.</p> |
| Teacher Education and Development Study in Mathematics, TEDS-M 2008 | <p>Brese, F., & Tatto, M.T. (Eds.). (2012). <i>TEDS-M 2008 user guide for the international database</i>. Amsterdam: IEA.</p> <p>Carnoy, M., Beteille, T., Brodziak, I., Loyalka, P., & Luschi, T. (2009). <i>Teacher Education and Development Study in Mathematics (TEDS-M): Do countries paying teachers higher relative salaries have higher student mathematics achievement?</i> Amsterdam: IEA.</p> <p>Ingvarson, L., Schwille, J., Tatto, M.T., Rowley, G., Peck, R., & Senk, S.L. (2013). <i>An analysis of teacher education context, structure, and quality-assurance arrangements in TEDS-M countries: Findings from the IEA Teacher Education and Development Study in Mathematics (TEDS-M)</i>. Amsterdam: IEA.</p> <p>Schwille, J., Ingvarson, L., & Holdgreve-Resendez, R. (Eds.). (2013). <i>TEDS-M encyclopedia: A guide to teacher education context, structure, and quality assurance in 17 countries. Findings from the IEA Teacher Education and Development Study in Mathematics (TEDS-M)</i>. Amsterdam: IEA.</p> <p>Tatto, M.T. (Ed.). (2013). <i>The Teacher Education and Development Study in Mathematics (TEDS-M): Policy, practice, and readiness to teach primary and secondary mathematics in 17</i></p> |

| | |
|--|---|
| | <p>countries. Technical report. Amsterdam: IEA.</p> <p>Tatto, M.T., Schwille, J., Senk, S.L., Ingvarson, L., Peck, R., & Rowley, G. (2008). <i>Teacher Education and Development Study in Mathematics (TEDS-M): Policy, practice, and readiness to teach primary and secondary mathematics. Conceptual framework</i>. Amsterdam: IEA.</p> <p>Tatto, M.T., Schwille, J., Senk, S.L., Ingvarson, L., Rowley, G., Peck, R., Bankov, K., Rodriguez, M., & Reckase, M. (2012). <i>Policy, practice, and readiness to teach primary and secondary mathematics in 17 countries: Findings from the IEA Teacher Education and Development Study in Mathematics (TEDS-M)</i>. Amsterdam: IEA.</p> |
| International Civic and Citizenship Education Study 2009, ICCS 2009 | <p>Ainley, J., Schulz, W., & Friedman, T. (Eds.). (2013). <i>ICCS 2009 encyclopedia: Approaches to civic and citizenship education around the world</i>. Amsterdam: IEA.</p> <p>Brese, F., Jung, M., Mirazchiyski, P., Schulz, W., & Zuehlke, O. (2014). <i>ICCS 2009 user guide for the international database (2nd ed.)</i>. Amsterdam: IEA.</p> <p>Fraillon, J., Schulz, W., & Ainley, J. (2012). <i>ICCS 2009 Asian report: Civic knowledge and attitudes among lower-secondary students in five Asian countries</i>. Amsterdam: IEA.</p> <p>Kerr, D., Sturman, L., Schulz, W., & Burge, B. (2010). <i>ICCS 2009 European report: Civic knowledge, attitudes, and engagement among lower-secondary students in 24 European countries</i>. Amsterdam: IEA.</p> <p>Schulz, W., Ainley, J., & Fraillon, J. (Eds.). (2011). <i>ICCS 2009 technical report</i>. Amsterdam: IEA.</p> <p>Schulz, W., Ainley, J., Fraillon, J., Kerr, D., & Losito, B. (2010). <i>ICCS 2009 international report: Civic knowledge, attitudes, and engagement among lower-secondary school students in 38 countries</i>. Amsterdam: IEA.</p> <p>Schulz, W., Ainley, J., Fraillon, J., Kerr, D., & Losito, B. (2010). <i>Initial findings from the IEA International Civic and Citizenship Education Study</i>. Amsterdam: IEA.</p> <p>Schulz, W., Ainley, J., Fraillon, J., Kerr, D., & Losito, B. (2010). <i>Prvotní zjištění z Mezinárodní studie občanské výchovy [Initial findings from the IEA International Civic and Citizenship Education Study]</i>. Prague: ÚIV.</p> <p>Schulz, W., Ainley, J., Fraillon, J., Kerr, D., & Losito, B. (2010). <i>Resultados iniciales del Estudio Internacional de Educación Cívica y Ciudadana de la IEA [Initial findings from the IEA International Civic and Citizenship Education Study]</i>. Amsterdam: IEA.</p> <p>Schulz, W., Ainley, J., Friedman, T., & Lietz, P. (2011). <i>ICCS 2009 Latin American report: Civic knowledge and attitudes among lower-secondary students in six Latin American countries</i>. Amsterdam: IEA.</p> <p>Schulz, W., Ainley, J., Friedman, T., & Lietz, P. (2011). <i>Informe Latinoamericano del ICCS 2009: Actitudes y conocimientos cívicos de estudiantes de secundaria en seis países de América Latina [ICCS 2009 Latin American report: Civic knowledge and attitudes among lower-secondary students in six Latin American countries]</i>. Amsterdam: IEA.</p> <p>Schulz, W., Fraillon, J., Ainley, J., Losito, B., & Kerr, D. (2008). <i>International Civic and Citizenship Education Study: Assessment framework</i>. Amsterdam: IEA.</p> <p>Schulz, W., Fraillon, J., Ainley, J., Losito, B., & Kerr, D. (2010). <i>Estudio internacional sobre educación cívica y ciudadana: Marco de la evaluación [International Civic and Citizenship Education Study: Assessment framework]</i>. Madrid: Instituto de Evaluación.</p> <p>Schulz, W., Fraillon, J., Ainley, J., Losito, B., & Kerr, D. (2010). <i>Indagine Internazionale sull'Educazione Civica e alla Cittadinanza: Quadro di riferimento [International Civic and Citizenship Education Study: Assessment framework]</i>. Frascati, Italy: INVALSI.</p> |
| Progress in International Reading Literacy Study 2011, PIRLS 2011 | <p>Foy, P. (2013). <i>TIMSS and PIRLS 2011 user guide for the fourth grade combined international database</i>. Chestnut Hill, MA: Boston College.</p> <p>Foy, P., & Drucker, K.T. (Eds.). (2013). <i>PIRLS 2011 user guide for the international database</i>. Chestnut Hill, MA: Boston College.</p> <p>Martin, M.O., & Mullis, I.V.S. (Eds.). (2012). <i>Methods and procedures in TIMSS and PIRLS 2011</i>. Chestnut Hill, MA: Boston College.</p> <p>Martin, M.O., & Mullis, I.V.S. (Eds.). (2013). <i>TIMSS and PIRLS 2011: Relationships among reading, mathematics, and science achievement at the fourth grade—implications for early learning</i>. Chestnut Hill, MA: Boston College.</p> <p>Mullis, I.V.S., Martin, M.O., Foy, P., & Drucker, K.T. (2012). <i>PIRLS 2011 international results in reading</i>. Chestnut Hill, MA: Boston College.</p> <p>Mullis, I.V.S., Martin, M.O., Kennedy, A.M., Trong, K.L., & Sainsbury, M. (2009). <i>PIRLS 2011 assessment framework</i>. Chestnut Hill, MA: Boston College.</p> <p>Mullis, I.V.S., Martin, M.O., Minnich, C.A., Drucker, K.T., & Ragan, M.A. (Eds.). (2012). <i>PIRLS 2011 encyclopedia: Education policy and curriculum in reading (Vols. 1–2)</i>. Chestnut Hill, MA: Boston College.</p> |
| Trends in International Mathematics and Science Study 2011, TIMSS 2011 | <p>Foy, P. (2013). <i>TIMSS and PIRLS 2011 user guide for the fourth grade combined international database</i>. Chestnut Hill, MA: Boston College.</p> <p>Foy, P., Arora, A., & Stanco, G.M. (Eds.). (2013). <i>TIMSS 2011 user guide for the international database</i>. Chestnut Hill, MA: Boston College.</p> |

| | |
|---|---|
| | <p>Martin, M.O., & Mullis, I.V.S. (Eds.). (2012). <i>Methods and procedures in TIMSS and PIRLS 2011</i>. Chestnut Hill, MA: Boston College.</p> <p>Martin, M.O., & Mullis, I.V.S. (Eds.). (2013). <i>TIMSS and PIRLS 2011: Relationships among reading, mathematics, and science achievement at the fourth grade—implications for early learning</i>. Chestnut Hill, MA: Boston College.</p> <p>Martin, M.O., Mullis, I.V.S., Foy, P., & Stanco, G.M. (2012). <i>TIMSS 2011 international results in science</i>. Chestnut Hill, MA: Boston College.</p> <p>Mullis, I.V.S., Martin, M.O., Foy, P., & Arora, A. (2012). <i>TIMSS 2011 international results in mathematics</i>. Chestnut Hill, MA: Boston College.</p> <p>Mullis, I.V.S., Martin, M.O., Minnich, C.A., Stanco, G.M., Arora, A., Centurino, V.A.S., & Castle, C.E. (Eds.). (2012). <i>TIMSS 2011 encyclopedia: Education policy and curriculum in mathematics and science (Vols. 1–2)</i>. Chestnut Hill, MA: Boston College.</p> <p>Mullis, I.V.S., Martin, M.O., Ruddock, G.J., O'Sullivan, C.Y., & Preuschoff, C. (2009). <i>TIMSS 2011 assessment frameworks</i>. Chestnut Hill, MA: Boston College.</p> <p>Mullis, I.V.S., Martin, M.O., Ruddock, G.J., O'Sullivan, C.Y., & Preuschoff, C. (2012). <i>Quadro di riferimento di TIMSS 2011 [TIMSS 2011 assessment frameworks]</i>. Frascati, Italy: INVALSI.</p> <p>Mullis, I.V.S., Martin, M.O., Ruddock, G.J., O'Sullivan, C.Y., & Preuschoff, C. (2012). <i>TIMSS 2011 Marcos de la evaluación [TIMSS 2011 assessment frameworks]</i>. Madrid: INEE.</p> |
| International Computer and Information Literacy Study, ICILS 2013 | Fraillon, J., Schulz, W., & Ainley, J. (2013). <i>International Computer and Information Literacy Study: Assessment framework</i> . Amsterdam: IEA. |
| Trends in International Mathematics and Science Study 2015, TIMSS 2015 | Mullis, I.V.S., & Martin, M.O. (Eds.). (2013). <i>TIMSS 2015 assessment frameworks</i> . Chestnut Hill, MA: Boston College. |
| Trends in International Mathematics and Science Study Advanced, TIMSS Advanced 2015 | Mullis, I.V.S., & Martin, M.O. (Eds.). (2014). <i>TIMSS Advanced 2015 assessment frameworks</i> . Chestnut Hill, MA: Boston College. |
| Progress in International Reading Literacy Study 2016, PIRLS 2016 | Mullis, I.V.S., & Martin, M.O. (Eds.). (2013). <i>PIRLS 2016 assessment framework</i> . Chestnut Hill, MA: Boston College. |
| International Civic and Citizenship Education Study 2016, ICCS 2016 | |
| Early Childhood Education Study, ECES | |

OECD Publications (PISA)

| | Publications with results and key findings |
|-----------|--|
| PISA 2000 | Knowledge and Skills for Life: First Results from the OECD Programme for International Student Assessment (PISA) 2000. OECD: Paris |
| PISA 2000 | Literacy Skills for the World of Tomorrow: Further Results from PISA 2000 - Publications 2000 http://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/literacyskillsfortheworldofomorrowfurtherresultsfrompisa2000-publications2000.htm |
| PISA 2003 | Learning for Tomorrow's World – First Results from PISA 2003 http://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/learningfortomorrowsworldfirstresultsfrompisa2003.htm |
| PISA 2006 | PISA 2006 Science Competencies for Tomorrow's World (Volume I and II) http://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/pisa2006results.htm |
| PISA 2009 | Volume I, What Students Know and Can Do: Student Performance in Reading, Mathematics and Science http://www.oecd.org/pisa/pisaproducts/pisa2009resultswatstudentsknowandcandostudentperformanceinreadingmathematicsandsciencevolumei.htm |
| PISA 2009 | Volume II, Overcoming Social Background: Equity in Learning Opportunities and Outcomes http://www.oecd.org/pisa/pisaproducts/pisa2009resultsovercomingsocialbackgroundequityinlearningopportunitiesandoutcomesvolumeii.htm |
| PISA 2009 | Volume III, Learning to Learn: Student Engagement, Strategies and Practices http://www.oecd.org/pisa/pisaproducts/pisa2009resultslearningtolearnvolumeiii.htm |
| PISA 2009 | Volume IV, What Makes a School Successful? Resources, Policies and Practices http://www.oecd.org/pisa/pisaproducts/pisa2009resultswatmakesaschoolsuccessfulvolumeiv.htm |
| PISA 2009 | Volume V., Learning Trends: Changes in Student Performance Since 2000 |

| Publications with results and key findings | |
|---|---|
| | http://www.oecd.org/pisa/pisaproducts/pisa2009resultslearningtrendschangesinstudentperformancesince2000volumev.htm |
| PISA 2009 | Volume VI, Students On Line: Digital Technologies and Performance http://www.oecd.org/pisa/pisaproducts/pisa2009resultsstudentsonlinedigitaltechnologiesandperformancevolumevi.htm |
| PISA 2012 | What Students Know and Can Do: Student Performance in Mathematics, Reading and Science (Volume I) http://www.oecd.org/pisa/keyfindings/pisa-2012-results-volume-i.htm |
| PISA 2012 | Excellence through Equity: Giving Every Student the Chance to Succeed (Volume II) http://www.oecd.org/pisa/keyfindings/pisa-2012-results-volume-ii.htm |
| PISA 2012 | Ready to Learn: Students' Engagement, Drive and Self-Beliefs (Volume III) http://www.oecd.org/pisa/keyfindings/pisa-2012-results-volume-iii.htm |
| PISA 2012 | What Makes Schools Successful? Resources, Policies and Practices (Volume IV) http://www.oecd.org/pisa/keyfindings/pisa-2012-results-volume-iv.htm |
| PISA 2012 | Creative Problem Solving: Students' skills in tackling real-life problems (Volume V) http://www.oecd.org/pisa/keyfindings/pisa-2012-results-volume-v.htm |
| PISA 2012 | Students and Money: Financial Literacy Skills for the 21st century (Volume VI) http://www.oecd.org/pisa/keyfindings/pisa-2012-results-volume-vi.htm |
| PISA in depth reports | |
| Reading | |
| PISA 2000 | Reading for change: performance and engagement across countries (PISA 2000) http://www.oecd.org/edu/school/programmeforminternationalstudentassessmentpisa/readingforchangeperformanceandengagementacrosscountries-publications2000.htm |
| | Let's Read Them a Story! The Parent Factor in Education http://www.oecd.org/edu/school/programmeforminternationalstudentassessmentpisa/pisa-letsreadthemastorytheparentfactorineducation.htm |
| Mathematics and Problem Solving | |
| PISA 2003 | Learning mathematics for life: A perspective from PISA (PISA 2003) http://www.oecd.org/edu/school/programmeforminternationalstudentassessmentpisa/learningmathematicsforlifea-perspectivefrompisa.htm |
| PISA 2003 | Mathematics teaching and learning strategies in PISA (PISA 2003) |
| PISA 2003 | Problem Solving for Tomorrow's World: First Measures of Cross-Curricular Competencies from PISA 2003 http://www.oecd.org/edu/school/programmeforminternationalstudentassessmentpisa/problemsolvingfortomorrowsworldfirstmeasuresofcrosscurricularcompetenciesfrompisa2003.htm |
| Science and Technology | |
| PISA 2006 | Green at fifteen? How 15-year-olds perform in environmental science and geoscience in PISA 2006 http://www.oecd.org/edu/school/programmeforminternationalstudentassessmentpisa/greenatfifteenhow15-year-oldsperforminenvironmentalscienceandgeoscienceinpisa2006.htm |
| PISA 2006 | Top of the class: high performers in science in PISA 2006 http://www.oecd.org/edu/school/programmeforminternationalstudentassessmentpisa/topoftheclass-highperformersinscienceinpisa2006.htm |
| PISA 2006 | PISA Computer-based assessment of science (PISA 2006) http://www.oecd.org/denmark/pisacomputer-basedassessmentofscience.htm |
| PISA 2006 | Are students ready for a technology-rich world? What PISA studies tell us (PISA 2003) http://www.oecd.org/edu/school/programmeforminternationalstudentassessmentpisa/arestudentsreadyforatechnology-richworldwhatpisastudiestellus.htm |
| Student learning | |
| PISA 2000 | Learners for life: student approaches to learning (PISA 2000) http://www.oecd.org/edu/school/programmeforminternationalstudentassessmentpisa/learnersforlifestudentapproachestolearning-publications2000.htm |
| PISA 2000 | Student engagement at school: a sense of belonging and participation (PISA 2000) http://www.oecd.org/edu/school/programmeforminternationalstudentassessmentpisa/studentengagementatschoolasenseofbelongingandparticipation-publicationspisa2000.htm |
| PISA 2006 | Quality time for students: learning in and out of school (PISA 2006) http://www.oecd.org/edu/school/programmeforminternationalstudentassessmentpisa/pisa-qualitytimeforstudentslearninginandoutofschool.htm |
| PISA 2009 | PISA - Grade Expectations: How Marks and Education Policies Shape Students' Ambitions (PISA 2009) http://www.oecd.org/pisa/pisaproducts/pisa-untappedskillsrealisingthepotentialofimmigrantstudents.htm |
| PISA 2009 | Learning to Learn: Student Engagement, Strategies and Practices (PISA 2009 results: Volume III) (this volume is also presented above as a result report) http://www.oecd.org/pisa/pisaproducts/pisa2009resultslearningtolearnvolumeiii.htm |
| Equity | |
| PISA 2000 | School factors related to quality and equity (PISA 2000) http://www.oecd.org/edu/school/programmeforminternationalstudentassessmentpisa/schoolfactorsrelatedtoquality |

| Publications with results and key findings | |
|---|--|
| | yandequity-pisa2000publications.htm |
| PISA 2000 | What makes school systems perform? (PISA 2000) http://www.oecd.org/edu/school/programme-for-international-student-assessment-pisa/what-makes-school-systems-perform-publications-pisa2000.htm |
| PISA 2003 | Where immigrant students succeed - A comparative review of performance and engagement in PISA 2003 |
| PISA 2006 | Equally prepared for life? How 15-year-old boys and girls perform in school (PISA 2006) http://www.oecd.org/edu/school/programme-for-international-student-assessment-pisa/equally-prepared-for-life-how-15-year-old-boys-and-girls-perform-in-school.htm |
| PISA 2009 | Overcoming Social Background: Equity in Learning Opportunities and Outcomes (PISA 2009 results: Volume II)(this volume is also presented above as a result report) http://www.oecd.org/pisa/pisaproducts/pisa2009resultsovercomingsocialbackgroundequityinlearningopportunitiesandoutcomesvolumeii.htm |
| Economics of Education | |
| PISA 2006 | The high cost of low educational performance (PISA 2006) http://www.oecd.org/edu/school/programme-for-international-student-assessment-pisa/the-high-cost-of-low-educational-performance.htm |
| | Public and Private Schools: How management and funding relate to their socio-economic profile http://www.oecd.org/edu/school/programme-for-international-student-assessment-pisa/public-and-private-schools-how-management-and-funding-relate-to-their-socio-economic-profile.htm |
| Technical reports | |
| PISA 2000 | PISA 2000 Technical Report, Edited by Ray Adams & Margaret Wu http://www.oecd.org/pisa/pisaproducts/33688233.pdf |
| PISA 2003 | PISA 2003 Technical Report http://www.oecd.org/pisa/pisaproducts/pisa2003technicalreport.htm |
| PISA 2006 | PISA 2006 Technical Report http://www.oecd.org/pisa/pisaproducts/42025182.pdf |
| PISA 2009 | PISA 2009 Technical Report http://www.oecd.org/pisa/pisaproducts/pisa2009technicalreport.htm |
| Assessment frameworks | |
| PISA 2000 | Measuring Student Knowledge and Skills: A New Framework for Assessment http://www.oecd.org/edu/school/programme-for-international-student-assessment-pisa/33693997.pdf |
| PISA 2003 | The PISA 2003 Assessment Framework – Mathematics, Reading, Science and Problem Solving Knowledge and Skills http://www.oecd.org/edu/school/programme-for-international-student-assessment-pisa/33694881.pdf |
| PISA 2006 | Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006 http://www.oecd-ilibrary.org/content/book/9789264026407-en |
| PISA 2009 | PISA 2009 Assessment Framework – Key Competencies in Reading, Mathematics and Science http://www.oecd.org/pisa/pisaproducts/pisa2009assessmentframework-keycompetenciesinreadingmathematicsandscience.htm |
| PISA 2012 | PISA 2012 Frameworks – Mathematics, Problem Solving and Financial Literacy http://www.oecd.org/pisa/pisaproducts/pisa2012draftframeworks-mathematicsproblemsolvingandfinancialliteracy.htm |
| PISA 2015 | PISA 2015 Draft Frameworks http://www.oecd.org/pisa/pisaproducts/pisa2015draftframeworks.htm |
| Research papers: OECD Education Working Papers (some of these papers are not exclusively discussing PISA). (Affiliation is just mentioned when it is other than OECD, Paris) | |
| No.1 2002 | Teacher Demand and Supply: Improving Teaching Quality and Addressing Teacher Shortages Paulo Santiago, (DOI 10.1787/232506301033) |
| No.2 2002 | Teacher Education and the Teaching Career in an Era of Lifelong Learning John Coolahan, (DOI 10.1787/226408628504) |
| No.3 2003 | The Role of National Qualifications Systems in Promoting Lifelong Learning Friederike Behringer, Mike Coles, (DOI 10.1787/224841854572) |
| No.4 2005 | Measuring Educational Productivity in Standards-Based Accountability Systems Martin Hampel, (DOI 10.1787/224417012465) |
| No.5 2006 | PISA 2000: Sample Weight Problems in Austria Erich Neuwirth (DOI 10.1787/220456725273) |
| No.6 2007 | Funding Systems and their Effects on Higher Education Systems Franz Strehl, Sabine Reisinger, Michael Kalatschan, (DOI 10.1787/220244801417) |
| No.7 2007 | On the Edge: Securing a Sustainable Future for Higher Education OECD, (DOI 10.1787/220180871707) |
| No.8 2007 | Globalisation and Higher Education Simon Marginson, Marijk van der Wende (University of Melbourne, Australia, University of Twente, Netherlands), (DOI 10.1787/173831738240) |

| | Publications with results and key findings |
|------------|---|
| No.9 2007 | Understanding the Regional Contribution of Higher Education Institutions: A Literature Review Peter Arbo & Paul Benneworth (University of Tromsø, Norway, University of Newcastle, UK (DOI 10.1787/161208155312) |
| No.10 2007 | Effects of Tertiary Expansion: Crowding-out effects and labour market matches for the higher educated Bo Hansson, (DOI 10.1787/085513474523) |
| No.11 2007 | Skilled Voices? Reflections on Political Participation and Education in Austria Florian Walter & Sieglinde Rosenberger (University of Economics and Business, Vienna, Austria) (DOI 10.1787/050662025403) |
| No.12 2007 | Education and Civic Engagement: Review of Research and a Study on Norwegian Youths Jon Lauglo & Tormod Øia (University of Oslo and Telemark University College, Prossgrunn, Norway) (DOI 10.1787/050484723262) |
| No.13 2007 | School Accountability, Autonomy, Choice, and the Level of Student Achievement: International Evidence from PISA 2003 Ludger Wöbmann, Elke Lüdemann, Gabriela Schütz & Martin R West (IFO Institute for Economic Research University in Munich, Germany and Brown University, Providence, RI, USA) (DOI 10.1787/246402531617) |
| No.14 2007 | School Accountability, Autonomy, Choice, and the Equity of Student Achievement: International Evidence from PISA 2003 Gabriela Schütz, Martin R West & Ludger Wöbmann (IFO Institute for Economic Research University in Munich, Germany and Brown University, Providence, RI, USA) (DOI 10.1787/246374511832) |
| No.15 2008 | Assessment of Learning Outcomes in Higher Education: a comparative review of selected practices Deborah Nusche (DOI 10.1787/244257272573) |
| No.16 | Missing on http://www.oecd-ilibrary.org/education/oecd-education-working-papers_19939019?page=2 |
| No.17 2009 | Recent Developments in Intellectual Capital Reporting and their Policy Implications W Richard Frederick (DOI 10.1787/227362757626) |
| No.18 2009 | Employers' Perspectives on the Roles of Human Capital Development and Management in Creating Value Bo Hansson (DOI 10.1787/227353534651) |
| No.19 2008 | Job-Related Training and Benefits for Individuals: A Review of Evidence and Explanations Bo Hansson (DOI 10.1787/237755412637) |
| No.20 2008 | A Framework for Monitoring Transition Systems Rolf K W van der Velden & Maarten H J Wolbers (University of Maastricht, Radboud University, Netherlands) (DOI 10.1787/221381866820) |
| No.21 2008 | Final Report of the Development of an International Adult Learning Module (OECD AL Module) Recommendations on Methods, Concepts and Questions in International Adult Learning Surveys Helmut Kuwan & Ann-Charlotte Larsson (Social Research and Consultancy, Germany and Statistics Sweden) (DOI 10.1787/236208471741) |
| No.22 2009 | What Works in Migrant Education? A Review of Evidence and Policy Options Deborah Nusche (DOI 10.1787/227131784531) |
| No.23 2009 | Teacher Evaluation: Current Practices in OECD Countries and a Literature Review Marlene Isoré (DOI 10.1787/223283631428) |
| No.24 2009 | Assessment and Innovation in Education Janet W Looney (DOI 10.1787/222814543073) |
| No.25 2009 | Do Quasi-markets Foster Innovation in Education? A Comparative Perspective Christopher Lubienski (University of Illinois, USA) (DOI 10.1787/221583463325) |
| No.26 2009 | International Adult Literacy and Basic Skills Surveys in the OECD Region William Thom (DOI 10.1787/221351213600) |
| No.27 | Missing on http://www.oecd-ilibrary.org/education/oecd-education-working-papers_19939019?page=2 |
| No.28 2010 | Les compétences bien sûr, mais pas seulement L'influence des compétences cognitives en lecture sur la participation aux études postsecondaires chez les jeunes Canadiens (DOI 10.1787/5kmlw23rnl-fr) |
| No.29 2010 | Impact of Proficiency on Early Entrants to the Labour Market: Evidence from the YITS Torben Drewes (Trent University, Canada) (DOI 10.1787/5kmlwvx7cr7-en) |
| No.30 2010 | How Does Academic Ability Affect Educational and Labour Market Pathways in Canada Jorgen Hansen (DOI 10.1787/5kmlw10hl6l-en) |
| No.31 2010 | Does Reading Proficiency at Age 15 Affect Pathways through Learning and Work Tomasz Gluszynski, Justin Bayard (DOI 10.1787/5kmlwvm3rbq-en) |
| No.32 2010 | Comparing the Similarities and Differences of PISA 2003 and TIMSS Margaret Wu (University of Melbourne, Australia) (DOI 10.1787/5km4psnm13nx-en) |
| No.33 2009 | PIAAC Reading Component: A Conceptual Framework John P Sabatini & Kelly M Bruce (Educational Testing Service, USA) (DOI 10.1787/220367414132) |
| No.34 2009 | PIAAC Literacy: A Conceptual Framework (DOI 10.1787/220348414075) |
| No.35 2009 | PIAAC Numeracy: A Conceptual Framework (DOI 10.1787/220337421165) |

| | Publications with results and key findings |
|------------|--|
| No.36 2009 | PIAAC Problem Solving in Technology-Rich Environments: A Conceptual Framework (DOI 10.1787/220262483674) |
| No.37 | Missing on http://www.oecd-ilibrary.org/education/oecd-education-working-papers_19939019?page=2 |
| No.38 2009 | ICT in Initial Teacher Training: Research Review Ann-Britt Enochsson & Caroline Rizza (Karlstad University, Sweden & Telecom, Paris, France) (DOI 10.1787/220502872611) |
| No.39 2009 | Education and Obesity in Four OECD Countries Franco Sassi, Marion Devaux, Jody Church, Michele Cecchini & Francesca Borgonovi (DOI 10.1787/5km4psmtn8zx-en) |
| No.40 | Missing on http://www.oecd-ilibrary.org/education/oecd-education-working-papers_19939019?page=2 |
| No.41 2009 | 21st Century Skills and Competences for New Millennium Learners in OECD Countries Katerina Ananiadou & Magdalen Claro (DOI 10.1787/218525261154) |
| No.42 2009 | School Evaluation: Current Practices in OECD Countries and a Literature Review Violaine Faubert (Institut d'Études Politiques, France) (DOI 10.1787/218816547156) |
| No.43 2010 | Summary of the June 2009 educationtoday Crisis Survey: Initial Reflections on the Impact of the Economic Crisis on Education Kiira Karkkainen (DOI 10.1787/5kmmsxbws77d-en) |
| No.44 2010 | 1-1 in Education: Current Practice, International Comparative Research Evidence and Policy Implications Oscar Valiente (DOI 10.1787/5kmjzwl9vr2-en) |
| No.45 2010 | How Technology Changes Demands for Human Skills Frank Levy (Massachusetts Institute of Technology, USA) (DOI 10.1787/5kmhds6czqzq-en) |
| No.46 2010 | Analysis of PISA 2006 Preferred Items Ranking Using the Percent-Correct Method Ray Adams, Alla Berezner & Maciej Jakubowski (Australian Council for Educational Research, Australia, Warsaw University, Poland) (DOI 10.1787/5km4psmntkq5-en) |
| No.47 2010 | International Approaches to Teacher Selection and Recruitment Andrew J Hobson, Patricia Ashby, Joanna McIntyre & Angi Malderez (University of Nottingham, UK) (DOI 10.1787/5kmbpnhh6qmx-en) |
| No.48 2010 | Initial Teacher Education and Continuing Training Policies in a Comparative Perspective: Current Practices in OECD Countries and a Literature Review on Potential Effects Pauline Musset (DOI 10.1787/5kmbpnhh7s47h-en) |
| No.49 2011 | The Impact of the 1999 Education Reform in Poland (DOI 10.1787/5kmbjgkm1m9x-en) |
| No.50 2010 | Education, Alcohol Use and Abuse Among Young Adults in Britain Maria del Carmen Huerta & Francesca Borgonovi (DOI 10.1787/5kmbqvsh57g0-en) |
| No.51 | Missing on http://www.oecd-ilibrary.org/education/oecd-education-working-papers_19939019?page=2 |
| No.52 2010 | Markets in Education: An Analytical Review of Empirical Research on Market Mechanisms in Education Sietske Waslander, Cissy Pater & Maartje van der Weide (University of Tilburg, Netherlands) (DOI 10.1787/5km4pskmkr27-en) |
| No.53 2010 | Taking on the Completion Challenge: A Literature Review on Policies to Prevent Dropout and Early School Leaving Cecilia S Lyche (DOI 10.1787/5km4m2t59cmr-en) |
| No.54 2010 | Using Student Test Results for Accountability and Improvement: A Literature Review Morten Anstorp Rosenkvist (Norwegian Ministry of Education and Research, Norway) (DOI 10.1787/5km4htwzvb30-en) |
| No.55 2011 | Workforce Skills and Innovation: An Overview of Major Themes in the Literature Philip Toner (University of Sydney, Australia) (DOI 10.1787/5kgk6hpnhxzq-en) |
| No.56 2011 | OECD Educationtoday Crisis Survey 2010: The Impact of the Economic Recession and Fiscal Crisis on Education in OECD Countries Dirk Van Damme & Kiira Karkkainen (DOI 10.1787/5kgj1r9zk09x-en) |
| No.57 2011 | Gendered Career Expectations of Students: Perspectives from PISA 2006 Joanna Sikora & Artur Pokropek (Australian National University, Australia and Polish Academy of Science, Poland) (DOI 10.1787/5kghw6891gms-en) |
| No.58 2011 | Integrating Formative and Summative Assessment: Progress Toward a Seamless System? Janet W Looney (DOI 10.1787/5kghx3kbl734-en) |
| No.59 2011 | Tuning-AHELO Conceptual Framework of Expected and Desired Learning Outcomes in Economics (DOI 10.1787/5kghtchwb3nn-en) |
| No.60 2011 | A Tuning-AHELO Conceptual Framework of Expected Desired/Learning Outcomes in Engineering (DOI 10.1787/5kghtchn8mbn-en) |
| No.61 2011 | ICT and Initial Teacher Education: National Policies Caroline Rizza (Telecom, Paris, France) (DOI 10.1787/5kg57kjj5hs8-en) |
| No.62 2012 | Access to Education Over the Working Life in Sweden: Priorities, Institutions and Efficiency Anders Stenberg (Swedish Institute for Social Research, Sweden) (DOI 10.1787/5k9fldmwwb-en) |
| No.63 2011 | An Analysis of Skill Mismatch Using Direct Measures of Skills Richard Desjardins & Kjell Rubensom (University of British Columbia, Canada) (DOI 10.1787/5kg3nh9h52g5- |

| | Publications with results and key findings |
|------------|--|
| | en) |
| No.64 2011 | Alignment in Complex Education Systems: Achieving Balance and Coherence Janet W Looney (DOI 10.1787/5kg3vg5lx8r8-en) |
| No.65 2011 | Student Standardised Testing: Current Practices in OECD Countries and a Literature Review Allison Morris (Institut d'Etudes Politiques, France) (DOI 10.1787/5kg3rp9qbnr6-en) |
| No.66 2012 | School Choice and Equity : Current Policies in OECD Countries and a Literature Review Pauline Musset (DOI 10.1787/5k9fq23507vc-en) |
| No.67 2012 | Exploring the Complex Interaction Between Governance and Knowledge in Education Mihály Fazekas & Tracey Burns (University of Cambridge, UK) (DOI 10.1787/5k9flcx2l340-en) |
| No.68 2012 | A Literature Review of School Practices to Overcome School Failure Brenton Faubert (Ontario Ministry of Education, Canada) (DOI 10.1787/5k9flcwv9tk-en) |
| No.69 2012 | Technical Feasibility of Reporting YTTs 2010 Skill Assessment Results on the PISA 2000 Reading Scale Fernando Cartwright (Statistics Canada, Canada) (DOI 10.1787/5k9fhndspvf1-en) |
| No.70 2012 | Guidelines for Quality Provision in Cross-Border Higher Education: Where Do We Stand? Stéphan Vincent-Lancrin & Sebastian Pfothenhauer (Massachusetts Institute of Technology, USA) (DOI 10.1787/5k9fd0kz0j6b-en) |
| No.71 2012 | The Policy Impact of PISA: An Exploration of the Normative Effects of International Benchmarking in School System Performance Simon Breakspear (University of Cambridge, UK) (DOI 10.1787/5k9fdqffr28-en) |
| No.72 2012 | Ageing and Skills: A Review and Analysis of Skill Gain and Skill Loss Over the Lifespan and Over Time Richard Desjardins & Arne Jonas Warnke (Centre for European Economics Research, Germany) (DOI 10.1787/5k9csvgw87ckh-en) |
| No.73 2012 | Parental Involvement in Selected PISA Countries and Economies Francesca Borgonovi & Guillermo Montt (DOI 10.1787/5k990rk0jsjj-en) |
| No.74 2012 | School Funding Formulas: Review of Main Characteristics and Impacts Mihály Fazekas (University of Cambridge, UK) (DOI 10.1787/5k993xw27cd3-en) |
| No.75 2012 | Immigrant Children's Age at Arrival and Assessment Results Anthony Heath & Elina Kilpi-Jakonen (University of Oxford, UK and Otto-Friedrich-University, Germany) (DOI 10.1787/5k993zsz6g7h-en) |
| No.76 2012 | Open Educational Resources: Analysis of Responses to the OECD Country Questionnaire Jan Hylén, Dirk Van Damme, Fred Mulder & Susan D'Antoni (Open Educational Resources, Sweden) (DOI 10.1787/5k990rjhvtlv-en) |
| No.77 2012 | Immigrant Status and Secondary School Performance as Determinants of Post-Secondary Participation: A Comparison of Canada and Switzerland Garnett Picot (Statistics Canada, Canada) (DOI 10.1787/5k9909jhz4w1-en) |
| No.78 2012 | Statistical Matching of PISA 2009 and TALIS 2008 Data in Iceland David Kaplan & Alyn Turner (University of Wisconsin, USA) (DOI 10.1787/5k97g3zzvg30-en) |
| No.79 2012 | Innovative Research-Based Approaches to Learning and Teaching Gesa Sonja Elsa van den Broek (Radboud University, Netherlands) (DOI 10.1787/5k97f6x1kn0w-en) |
| No.80 2012 | Subsidies and Levies as Policy Instruments to Encourage Employer-Provided Training Normann Müller & Friedrike Behringer (Bundesinstitut für Berufsbildung, Germany) (DOI 10.1787/5k97b083v1vb-en) |
| No.81 2012 | Hybrid Learning Environments: Merging Learning and Work Processes to Facilitate Knowledge Integration and Transitions Ilya Zitter & Aimée Hoeve (ECBO, Netherlands) (DOI 10.1787/5k97785xwvdf-en) |
| No.82 2012 | Bringing About Curriculum Innovations: Implicit Approaches in the OECD Area Kiira Kärkkäinen (DOI 10.1787/5k95qw8xzl8s-en) |
| No.83 2012 | Assessment for Qualification and Certification in Upper Secondary Education: A Review of Country Practices and Research Evidence Stefanie Dufaux (DOI 10.1787/5k92zp1cshvb-en) |
| No.84 2012 | Business-Driven Innovation: Is it Making a Difference in Education?: An Analysis of Educational Patents Dominique Foray & Julio Raffo (École Polytechnique Fédérale de Lausanne, Switzerland) (DOI 10.1787/5k91dl7pc835-en) |
| No.85 2012 | Looking Beyond the Numbers: Stakeholders and Multiple School Accountability Edith Hooge, Tracey Burns & Harald Wilkoszewski (Tilburg University, Netherlands) (DOI 10.1787/5k91dl7ct6q6-en) |
| No.86 2013 | Progression in Student Creativity in School: First Steps Towards New Forms of Formative Assessments Bill Lucas, Guy Claxton & Ellen Spencer (University of Winchester, UK) (DOI 10.1787/5k4dp59msdsk-en) |
| No.87 2013 | Analysis of the Predictive Power of PISA Test Items Maciej Jakubowski (Warsaw University, Poland) (DOI 10.1787/5k4bx47268g5-en) |
| No.88 2013 | Monitoring Adult Learning Policies: A Theoretical Framework and Indicators Anna Borkowsky (Sozialwissenschaftliche Forschung und Beratung, Switzerland) (DOI 10.1787/5k4c0vxjlkzt-en) |
| No.89 2013 | Learning from International Experiences with Interactive Whiteboards: The Role of Professional Development in Integrating the Technology |

| | Publications with results and key findings |
|-------------|---|
| No.90 2013 | Sara Hennessy & Laura London (University of Cambridge, UK) (DOI 10.1787/5k49chbsnmls-en) Review of the Italian Strategy for Digital Schools Francesco Awisati, Sara Hennessy, Robert B Kozma & Stéphan Vincent-Lancrin (University of Cambridge, UK) (DOI 10.1787/5k487ntdbr44-en) |
| No.91 2013 | Sparkling Innovation in STEM Education with Technology and Collaboration: A Case Study of the HP Catalyst Initiative Kiira Kärkkäinen & Stéphan Vincent-Lancrin (DOI 10.1787/5k480sj9k442-en) |
| No.92 2013 | Understanding School Building Policy and Practice in Belgium's Flemish Community Geert Leemans & Hannah von Ahlefeld (Flemish Agency for Infrastructure in Education, Belgium) (DOI 10.1787/5k46h2rtw5mx-en) |
| No.93 2013 | PISA in Low and Middle Income Countries Simone Bloem (DOI 10.1787/5k41tm2gx2vd-en) |
| No.94 2013 | Learning Study: Its Origins, Operationalisation, and Implications Eric C K Cheng & Mun Ling Lo (Hong Kong Institute of Education, China) (DOI 10.1787/5k3wj0s959p-en) |
| No.95 2014 | Schooling Matters: Opportunity to Learn in PISA 2012 William H Schmidt, Pablo Zoido & Leland Cogan (Michigan State University, USA) (DOI 10.1787/5k3v0hldmchl-en) |
| No.96 2013 | The Simple, the Complicated, and the Complex: Educational Reform Through the Lens of Complexity Theory Sean Snyder (University of Pennsylvania, USA) (DOI 10.1787/5k3txnpt1lnr-en) |
| No.97 2013 | Balancing Trust and Accountability? The Assessment for Learning Programme in Norway: A Governing Complex Education Systems Case Study Therese Hopfenbeck, Astrid Tolo, Teresa Florez & Yasmine El Masri (Oxford University, UK and University of Bergen, Norway) (DOI 10.1787/5k3txnpqlsnn-en) |
| No.98 2013 | Coping with Very Weak Primary Schools: Towards Smart Interventions in Dutch Education Policy Mark van Twist, Martijn van der Steen, MArieke Kleiboer, Jorren Scherpenisse & Hanno Theisens (Netherlands School of Public Administration, Netherlands, The Hague University for Applied Science, Netherlands) (DOI 10.1787/5k3txnpnhld7-en) |
| No.99 2013 | Learning Standards, Teaching Standards and Standards for School Principals: A Comparative Study Centre of Study for Policies and Practices in Education (CEPPE), Chile (DOI 10.1787/5k3tsjqtp90v-en) |
| No.100 2014 | Promoting Skills for Innovation in Higher Education: A Literature Review on the Effectiveness of Problem-based Learning and of Teaching Behaviours Sabine Hoidn & Kiira Kärkkäinen (Harvard University, USA) (DOI 10.1787/5k3tsj67l226-en) |
| No.101 2013 | Returns to Skills Around the World: Evidence from PIAAC Eric A Hanushek, Guido Schwerdt, Simon Wiederhold & Ludger Woessmann (Hoover Institution, Stanford University, USA, University of Siegen, Germany and Ifo Institute for Economic Research University of Munich, Germany) (DOI 10.1787/5k3tsjqmvtq2-en) |
| No.102 2014 | Comparison of PIAAC and PISA Frameworks for Numeracy and Mathematical Literacy Iddo Gal & Dave Tout (University of Haifa, Israel and Australian Council for Educational Research, Australia) (DOI 10.1787/5jz3wl63cs6f-en) |
| No.103 2014 | Evaluating Measurement Invariance of TALIS 2013 Complex Scales: Comparison between Continuous and Categorical Multiple-Group Confirmatory Factor Analyses Deana Desa (DOI 10.1787/5jz2kbbvlb7k-en) |
| No.104 2014 | Shifting Responsibilities - 20 Years of Education Devolution in Sweden: A Governing Complex Education Systems Case Study Patrick Blanchenay, Tracey Burns & Florian Köster (DOI 10.1787/5jz2jg1rqrd7-en) |
| | PISA in Focus articles |
| No.1 | Pre-primary education and performance in PISA (article not available online) |
| No.2 | Improving performance: leading from the bottom http://www.oecd.org/portugal/47271471.pdf |
| No.3 | Does investing in after-school classes pay off? http://www.oecd.org/portugal/47573005.pdf |
| No.4 | Has discipline in school deteriorated? http://www.oecd.org/pisa/pisaproducts/pisainfocus/47944912.pdf |
| No.5 | How do some students overcome their socio-economic background? http://www.oecd.org/pisa/pisaproducts/pisainfocus/48165173.pdf |
| No.6 | When students repeat grades or are transferred out of school: What does it mean for education systems? http://www.oecd.org/pisa/pisaproducts/pisainfocus/48363440.pdf |
| No.7 | Private schools: Who benefits? http://www.oecd.org/pisa/pisaproducts/pisainfocus/48482894.pdf |
| No.8 | Do students today read for pleasure? http://www.oecd.org/pisa/pisaproducts/pisainfocus/48624701.pdf |
| No.9 | School autonomy and accountability: Are they related to student performance? http://www.oecd.org/pisa/pisaproducts/pisainfocus/48910490.pdf |
| No.10 | What can parents do to help their children succeed in school? |

| | Publications with results and key findings |
|-------|--|
| | http://www.oecd.org/pisa/49012097.pdf |
| No.11 | How are school systems adapting to increasing numbers of immigrants students? http://www.oecd.org/pisa/pisaproducts/pisainfocus/49264831.pdf |
| No.12 | Are boys and girls ready for the digital age? http://www.oecd.org/pisa/pisaproducts/pisainfocus/49442737.pdf |
| No.13 | Does money buy strong performance in PISA? http://www.oecd.org/pisa/pisaproducts/pisainfocus/49685503.pdf |
| No.14 | What kinds of careers do boys and girls expect for themselves? http://www.oecd.org/pisa/pisaproducts/pisainfocus/49829595.pdf |
| No.15 | How green are today's 15-year-olds? http://www.oecd.org/pisa/pisaproducts/pisainfocus/50150271.pdf |
| No.16 | Does performance-based pay improve teaching? http://www.oecd.org/pisa/pisaproducts/pisainfocus/50328990.pdf |
| No.17 | Are large cities educational assets or liabilities? http://www.oecd.org/pisa/pisaproducts/pisainfocus/pisa%20in%20focus%20n17%20%28eng%29%20FINAL.pdf |
| No.18 | Are students more engaged when schools offer extracurricular activities? http://www.oecd.org/edu/pisa%20in%20focus%20n18%20%28eng%29--v05.pdf |
| No.19 | Is there really such a thing as a “second chance” in education? http://www.oecd.org/pisa/pisaproducts/pisainfocus/pisa%20in%20focus%20n19%20%28eng%29--v04--FINAL.pdf |
| No.20 | Are school vouchers associated with equity in education? http://www.oecd.org/pisa/pisaproducts/pisainfocus/pisa%20in%20focus%20n%C2%B020%20%28eng%29--v04_FINAL.pdf |
| No.21 | Do today's 15-year-olds feel environmentally responsible? http://www.oecd.org/pisa/pisaproducts/pisainfocus/pisa%20in%20focus%20n21%20%28eng%29--v04.pdf |
| No.22 | How do immigrant students fare in disadvantaged schools? http://www.oecd.org/pisa/pisaproducts/pisainfocus/pisa%20in%20focus%20n%C2%B022%20%28eng%29--Final%20bis.pdf |
| No.23 | What do students expect to do after finishing upper secondary school http://www.oecd.org/pisa/pisaproducts/pisainfocus/PIF23.pdf.pdf |
| No.24 | What do students think about school? http://www.oecd.org/pisa/pisaproducts/pisainfocus/pisa%20in%20focus%20n24%20%28eng%29--FINAL.pdf |
| No.25 | Are countries moving towards more equitable education systems? http://www.oecd.org/pisa/pisaproducts/pisainfocus/pisa%20in%20focus%20n25%20%28eng%29--FINAL.pdf |
| No.26 | Grade Expectations http://www.oecd.org/pisa/pisaproducts/pisainfocus/pisa%20in%20focus%20n26%20%28eng%29--FINAL.pdf |
| No.27 | Does it matter which school a student attends? http://www.oecd.org/pisa/pisaproducts/pisainfocus/pisa%20in%20focus%20n27%20%28eng%29--FINAL_version2.pdf |
| No.28 | What makes urban schools different? http://www.oecd.org/pisa/pisaproducts/pisainfocus/pisa%20in%20focus%20n28%20%28eng%29--FINAL.pdf |
| No.29 | Do immigrant students' reading skills depend on how long they've been in their new country? http://www.oecd.org/pisa/pisaproducts/pisainfocus/pisa%20in%20focus%20n29%20%28eng%29--Final.pdf |
| No.30 | Could learning strategies reduce the performance gap between advantaged and disadvantaged students? http://www.oecd.org/pisa/pisaproducts/pisainfocus/pisa%20in%20focus%20n30%20%28eng%29--Final.pdf |
| No.31 | Who are the academic all-rounders? http://www.oecd.org/pisa/pisaproducts/pisainfocus/pisa%20in%20focus%20n31%20%28eng%29--FINAL.pdf |
| No.32 | Do students perform better in schools with orderly classrooms? http://www.oecd.org/pisa/pisaproducts/pisainfocus/pisa%20in%20focus%20n32%20%28eng%29--v07.pdf |
| No.33 | What can immigrant students tell us about education systems? http://www.oecd.org/pisa/pisaproducts/pisainfocus/pisa-in-focus-n33-%28eng%29-FINAL.pdf |
| No.34 | Who are the strong performers and successful reformers in education? http://www.oecd.org/pisa/pisaproducts/pisainfocus/pisa-in-focus-n34-%28eng%29-FINAL.pdf |
| No.35 | Who are the school truants? http://www.oecd.org/pisa/pisaproducts/pisainfocus/PISA-in-Focus-n35-%28eng%29-FINAL.pdf |
| No.36 | Do parents' occupations have an impact on student performance? http://www.oecd.org/pisa/pisaproducts/pisainfocus/PISA-in-Focus-N36-%28eng%29-FINAL.pdf |
| No.37 | Do students have the drive to succeed? http://www.oecd.org/pisa/pisaproducts/pisainfocus/pisa-in-focus-n37-%28eng%29-final.pdf |
| No.38 | Are 15-year-olds creative problem-solvers? http://www.oecd.org/pisa/pisaproducts/pisainfocus/pisa-in-focus-n38-%28eng%29-final.pdf |
| No.39 | Are grouping and selecting students for different schools related to students' motivation to learn? |

| | Publications with results and key findings |
|-------|---|
| | http://www.oecd.org/pisa/pisaproducts/pisainfocus/pisa-in-focus-n39-%28eng%29-final.pdf |
| No.40 | Does pre-primary education reach those who need it most? http://www.oecd.org/pisa/pisaproducts/pisainfocus/pisa-in-focus-n40-%28eng%29-final.pdf |
| No.41 | Do 15-year-olds know how to manage money? http://www.oecd.org/pisa/pisaproducts/pisainfocus/pisa-in-focus-n41-%28eng%29-final.pdf |
| No.42 | When is competition between schools beneficial? http://www.oecd.org/pisa/pisaproducts/pisainfocus/PISA-in-Focus-N42-%28eng%29-FINAL.pdf |
| No.43 | Are disadvantaged students more likely to repeat grades? http://www.oecd.org/pisa/pisaproducts/pisainfocus/pisa-in-focus-n43-%28eng%29-final.pdf |
| No.44 | How is equity in resource allocation related to student performance? http://www.oecd.org/pisa/pisaproducts/pisainfocus/pisa-in-focus-n44-%28eng%29-final.pdf |

APPENDIX III: SHEET FOR CODING ARTICLES

| | |
|---|---|
| Title | |
| Author(s) | |
| Journal etc | |
| Abstract | |
| Keywords | |
| Aspects | |
| Which initiators | A. own B. Administration/Policy C. Other organization |
| Which countries | A. one country B. two or more |
| Explanandum (a few words to be categorized) | |
| What to be inquired? | Student achievement School characters Education system characters |
| How to be inquired? | Formal (only PISA) Formal plus formal (PISA + other LSA) Inside – outside (using PISA for inquiring other social phenomena) and substantial |
| What research question level? | Individual Organisational System Subject-specific |
| Type of account | Descriptive Explorative – hypothesis-generating Hypothesis-testing |
| What results? | |
| Which object of study (e.g. 8-graders in Shanghai and Sweden or high-school students in USA) | (If something extra) |
| What object of knowledge (e.g. relations between gender and self-concept in math achievement) | |
| What relevance – for whom? | Professional relevance Policy – administrative relevance Academic Relevance Citizen relevance General information |
| What is the logic from research to practical relevance? | Directives – how to improve achievements? Insight – how to understand achievements? Not related |
| Others | |

The Swedish Research Council carried out a project in 2014 in Educational Research, called SKOLFORSK. The work was commissioned by the government to survey Swedish and international research that is relevant to education. The aim was to create a platform of knowledge for the newly-formed Swedish Centre for Educational Research (Skolforskningsinstitutet). The conclusions of this report are those of the authors. The summary report by the Swedish Research Council, "Research and Education in collaboration" with a description of the project and the issues, results and recommendations presented in the sub-projects, as well as the other reports, can be downloaded from the Swedish Research Council website vr.se

