Science Teaching Practices and Learning Outcomes in High-Performing Asian Countries: Comparing TIMSS and PISA 2015

Anindito Aditomo

Faculty of Psychology, University of Surabaya, Indonesia

DIPF Leibniz Institute for Research and Information in Education, Germany

Abstract. This study explores the extent to which TIMSS and PISA yield consistent patterns of associations between science teaching and affective as well as cognitive outcomes in 5 high-performing Asian countries. Items with similar content were identified and selected to represent two forms of instruction: "inquiry" and "interactive" science teaching. Multi-group ESEM of the measurement model established metric-level invariance across the 5 countries. Subsequently, multi-group multilevel SEM revealed markedly different patterns of teaching-learning relationships between the two assessments. The results show that PISA and TIMSS do not always converge or lead to consistent conclusions, especially regarding the associations between teaching and cognitive outcome (achievement). Results from PISA indicate that inquiry is associated with lower achievement, while interactive teaching with higher achievement, In contrast, the weak and non-significant relations found in TIMSS suggest that the relations between both forms of teaching and learning outcomes may be more contextual. A better understanding of the reasons behind these diverging findings is necessary before stronger conclusions can be made about the effects of interactive and inquiry-based science teaching. This paper argues that one of the more probable explanations is related to the source of data (student perceptions vs. teacher reports). That is, teachers and students may interpret descriptions of teaching practices (e.g. "students are asked to draw conclusions from experiments") in systematically different ways. Further research is needed to test this and other possible explanations.

Keywords: science education, inquiry-based learning, intrinsic motivation, achievement, large-scale assessment

INTRODUCTION

This paper examines associations between science learning outcomes and two important forms of instruction: interactive teaching and inquiry-based teaching. Prior studies addressing this question have found that, contrary to expectations, inquiry is associated with poorer learning. Complementing prior studies which have mostly relied upon the PISA database, this paper draws upon TIMSS-Grade 8 and PISA 2015 to explore the extent to which consistent, meaningful patterns of findings could be observed in five high-performing Asian countries.

Teaching quality

Teaching is among the most important factors influencing student learning and achievement. Decades of research have attempted to identify characteristics of high quality teaching. While terminologies vary, these characteristics could be arranged under several broad dimensions (Fauth, Decristan, Rieser, Klieme, & Büttner, 2014; Neumann, Kauertz, & Fischer, 2012). One dimension is related to presentation, e.g. clear goals, logical sequence, and using appropriate pace. Another dimension is related to classroom climate, e.g. positive social interactions, orderly climate, and efficient transitions.

A third dimension, which is the focus of this paper, is often called cognitive activation (Klieme, Pauli, & Reusser, 2009).

Cognitive activation are practices which activate students' higher-order cognitive processes which are central to learning (Bransford, Brown, & Cocking, 2000). Cognitive activation practices can be generic or domain-general, such as questioning, prompting self-explanations, connecting new information with students' prior knowledge. They can also be domain-specific, such as inquiry activities often incorporated in science teaching. Inquiry is seen as essential for science learning when the goal is to develop not only conceptual knowledge, but also understanding about the nature of science.

Inquiry-based teaching

While widely endorsed, inquiry-based teaching has also attracted criticisms. Authors have argued that inquiry and other constructivist teaching approaches are often too unstructured, thus potentially overloading students' cognitive capacity and hampering learning (Kirschner, Sweller, & Clark, 2006; Mayer, 2004). Instead of student-driven approaches, these authors advocate for the use of explicit guidance and structure in teaching (Stockard, Wood, Coughlin, & Rasplica Khoury, 2018).

Conceptually, inquiry approaches do not necessarily exclude scaffolding or teacher guidance. Indeed, much research in science education have sought to investigate the nature and effects of various types of structures for the process and outcomes of inquiry activities (e.g. Quintana, Zhang, & Krajcik, 2005; Wilson, Taylor, Kowalski, & Carlson, 2010). Such studies have generally shown that structure and guidance increases the effectiveness of inquiry (Furtak, Seidel, Iverson, & Briggs, 2016; Lazonder & Harmsen, 2016).

The bulk of the literature on inquiry science teaching investigates innovative, research-based curricula implemented in specific classrooms/schools which may not be representative of more "naturally occurring" practices in the average classroom. In this regard, studies based on international large-scale assessment (ILSA) data provide an important contribution by examining teaching practices in a much wider range of schools. In general, ILSA studies have either found non-significant or negative associations between inquiry and science achievement (Areepattamannil, 2012; Cairns & Areepattamannil, 2017; Lay, Areepattamannil, Ng, & Khoo, 2015). However, other studies indicate that some items or dimensions of inquiry positively predict learning outcomes, at least in certain countries (Gee & Wong, 2012).

Science teaching in ILSA

TIMSS and PISA background questionnaires include items/indicators of teaching which reflect the concept of cognitive activation. These include domain-general practices such as asking students to explain their ideas and relating content to everyday life and students' prior knowledge. Others are more domain-specific, which in science refers to inquiry activities. These include practices such as asking students to design and conduct experiments, interpret data, and use evidence to support claims/conclusions. Comparatively, TIMSS covers more indicators of domain-general cognitive activation, while PISA includes more items on inquiry (Müller, Prenzel, Seidel, Schiepe-tiska, & Kjærnsli, 2016).

Beyond the specific items, the main difference between TIMSS' and PISA's assessment of teaching is their data sources. TIMSS uses teachers' report, while PISA assesses students' perceptions. Each approach has its strengths and drawbacks, and are often seen as complementary approaches to assess teaching quality (Lüdtke, Robitzsch, Trautwein, & Kunter, 2009).

Research question

To what extent do analyses based on TIMSS and PISA data produce consistent findings regarding the relationships between teaching practices and educational outcomes (enjoyment and achievement)?

METHOD

Data sources

Data was taken from the 2015 cycle of PISA and Grade-8 TIMSS from five high-performing Asian countries/education systems (Hong Kong, Japan, Korea, Singapore, and Taiwan). Teacher and student data from TIMSS were merged using the IDB Analyzer software.

Table 1. Number of students and class/schools.

ILSA	PISA		TIMSS	
Unit	*Students*	*Schools*	*Students*	*Schools*
Taiwan	7708	214	5964	208
Hong Kong	5359	138	4283	152
Japan	6647	198	5534	171
Korea	5581	168	6847	220
Singapore	6115	177	6116	334
Total	31410	895	28744	1085

Measures

In TIMSS, grade 8 science achievement is measured in terms students' knowledge in biology, chemistry, physics and earth science, and their ability to apply and reason using that knowledge (Jones, Wheeler, & Centurino, 2015). Similarly, PISA measures students' ability to explain natural phenomena, evaluate and design science investigations, and interpret scientific data and evidence (OECD, 2016). The key difference is that TIMSS' test content is curriculum-based, while PISA is not.

Enjoyment of learning science was chosen to represent affective outcome. This was measured by the IRT-scaled indices of "Students Like Learning Science" (TIMSS) and "Enjoyment of Science" (PISA).

Table 2 displays the indicators of teaching as cognitive activation used in this paper. Both TIMSS and PISA include several other indicators of cognitive activation (e.g. in PISA there are items referring to debate and argumentation around inquiry). However, only similarly-worded items considered to measure closely related constructs are chosen for the current analysis. In TIMSS, the data represent teacher reports, while in PISA they represent student perceptions.

Table 2. Indicators of inquiry and interactive science teaching practices

TIMSS	PISA
Inquiry-based teaching	
BTBG18E. Conduct experiments or investigations.	ST098Q02TA. Students spend time in the laboratory doing practical experiments.
BTBG18G. Interpret data from experiments or investigations.	ST098Q05TA. Students are asked to draw conclusions from an experiment they have conducted.
BTBG18D. Design or plan experiments or investigations.	ST098Q07TA. Students are allowed to design their own experiments.

Interactive teaching	
BTBG14B. Ask students to explain their answers.	ST098Q01TA. Students are given opportunities to explain their ideas.
BTBG14E. Link new content to students' prior knowledge.	ST098Q06TA. The teacher explains <school science> idea can be applied
BTBG14A. Relate the lesson to students' daily lives.	ST098Q09TA. The teacher clearly explains relevance <broad science> concepts to our lives.

Analysis

Analyses were conducted separately for TIMSS and PISA. Multi-group ESEM was applied using MPLUS to check for configural, metric, and scalar invariance for the two-factor model (inquiry vs. interactive teaching). Subsequently, multi-group multilevel SEM using MPLUS was applied to examine relations between teaching and learning outcomes (enjoyment and achievement) in three steps. First, a null model was fitted to estimate Level-1 (student) and Level-2 (classrooms in TIMSS and schools in PISA) variances of the dependent variables (enjoyment and achievement). Second, a random intercept model was fitted with gender (L1), immigrant status (L1), language spoken at home (L1), and SES (L1 and L2) as observed predictors. Third, interactive teaching and inquiry teaching, each measured by 3 items, were added as latent predictors. All available plausible values were utilised through the TYPE=IMPUTATION option. Design effects were taken into consideration by using the school id (PISA) and teacher-link (TIMSS) as clustering variables. The final student weight (PISA) and science teacher weight (TIMSS) variables were used to account for sampling bias. Due to the age-based sampling in PISA, students from a school might be taught by different teachers (there are no meaningful class-level cluster). Hence, teaching practices were treated as a student-level construct.

RESULTS

Preliminary results

Multi-group ESEM indicated that MI for the two factor model (interactive and inquiry teaching) could be established at the metric level (for PISA: CFI=.974, TLI=.963; RMSEA=.059; SRMR=.044; for TIMSS: CFI=.968, TLI=.953; RMSEA=.011; SRMR=.071). This implies that measurement models have the same number of factors and factor loadings across the 5 countries, allowing for meaningful comparison of regression slopes.

The multi-group multilevel SEM of the null models reveal a large variation of ICC for achievement: estimates were 25-44% (PISA) and 6-76% (TIMSS). For enjoyment, ICC were lower and less varied: between 1-10% (PISA) and 6-8% (TIMSS).

Teaching practices and student outcomes

To conserve space, only standardized slope estimates of teaching effects are reported. Among the covariates, individual and collective SES (i.e. ESCS in PISA and HLR in TIMSS) displayed consistently positive associations with outcomes.

Table 3. Effects of INQUIRY TEACHING (figures are standardised estimates and standard errors).

Outcome	Enjoyment		Achievement	
ILSA	*PISA*	*TIMSS*	*PISA*	*TIMSS*
Taiwan	0.00 (0.02)	0.05 (0.11)	**-0.47 (0.02) ****	**0.05 (0.05)**

Hong Kong	0.11 (0.04)**	0.37 (0.17)*	**-0.30 (0.05) **	**0.35 (0.11)**
Japan	-0.10 (0.02) **	0.10 (0.24)	-0.27 (0.03) **	0.03 (0.14)
Korea	-0.06 (0.02) **	-0.09 (0.12)	**-0.44 (0.02) **	**0.09 (0.08)**
Singapore	-0.01 (0.03)	0.05 (0.08)	-0.03 (0.04)	0.03 (0.05)

Table 4. Effects of INTERACTIVE TEACHING (figures are standardised estimates and standard errors).

Outcome	Enjoyment		Achievement	
ILSA	*PISA*	*TIMSS*	*PISA*	*TIMSS*
Taiwan	0.21 (0.02)**	0.08 (0.11)	**0.44 (0.02)**	**-0.01 (0.06)**
Hong Kong	0.27 (0.04)**	-0.09 (0.19)	**0.36 (0.05)**	**-0.10 (0.12)**
Japan	0.31 (0.02)**	-0.00 (0.22)	0.15 (0.03)**	-0.09 (0.13)
Korea	0.24 (0.02)**	0.36 (0.12)**	**0.26 (0.02)**	**0.00 (0.10)**
Singapore	0.32 (0.03)**	0.16 (0.10)	0.12 (0.04)**	0.02 (0.05)

DISCUSSION AND CONCLUSION

Estimates based on PISA indicate that inquiry is unrelated (or weakly related) to enjoyment and associated with lower achievement, but that interactive teaching is associated with both higher enjoyment and higher achievement. Meanwhile, estimates based on TIMSS mostly indicate both forms of teaching to be unrelated with enjoyment and achievement. For TIMSS, the only significant correlations were between inquiry and both outcomes in Hong Kong, and between interactive teaching and enjoyment in Korea.

At first blush, these results reveal little consistency between TIMSS and PISA. However, comparisons across PISA and TIMSS need to take into account the uncertainty (standard errors) around the point estimates. The standard errors in TIMSS are much larger because analyses were conduceted at the classroom level (i.e., much smaller sample size than PISA, which took students as its unit of analysis). Thus, for the purpose of comparing PISA and TIMSS, it is more appropriate to examine confidence intervals around the point estimates (and disregarding the statistical significance).

Using this approach, the discrepancy between PISA and TIMSS largely disappears with regards to the association between teaching and enjoyment. That is, there is too much uncertainty around the estimates to confidently infer that the findings reflect genuine differences. With regards to achievement, however, estimates from PISA and TIMSS seem to genuinely differ in Taiwan, Hong Kong, and Korea. For these three countries, PISA indicate that inquiry is associated with lower achievement, while interactive teaching with higher achievement.

Taken together, analyses based on PISA and TIMSS data do not always converge or lead to consistent conclusions. Inconsistencies between the two are more pronounced with regards to the association between teaching and cognitive outcomes. Results from PISA shed an unfavourable light on inquiry. Meanwhile, the weak and non-significant relations found in TIMSS suggest that inquiry's impact on learning may be more contextual. Thus, rather than leading to an unfavourable conclusion regarding inquiry, results from TIMSS would prompt further investigations of contextual features (e.g. teacher qualifications, school resources) which may moderate the relations between teaching and learning.

Which conclusions are more warranted? Answering this question would require a better understanding of why the diverging results emerged. Part of the explanation likely involves some aspect of the different methods adopted by TIMSS and PISA. One possibility stems from how

achievement is assessed (curriculum-based in TIMSS vs. daily life applications in PISA). However, the two assessments share much in common (Wu, 2009) and are highly correlated (Asendorpf & Conner, 2012). Furthermore, one might argue that because interactive teaching is teacher-centred, it should be linked more strongly with assessments which are curriculum-based (which was not the case here).

A more probable explanation might have to do with the source of data to assess teaching (students vs. teachers). In this case, the diverging results may reflect systematic differences in how a description such as "asking students to interpret data" might be interpreted. For most students, it may mean doing the activity without the teachers' guidance. For teachers, however, the same description may conflate guided and unguided activities (and hence its correlation with outcomes would depend upon context, as current results indicate).

This latter possibility is supported by a recent analysis of the TIMSS 2007 data for Taiwan, in which teaching was assessed based on students' perceptions but then aggregated at the class level (Liou & Jessie Ho, 2018). These authors found that science achievement is negatively related with inquiry, but positively with teacher-centred instruction. This is the same pattern of results obtained in this study based on the PISA data.

These possibilities are offered as conjectures to be systematically tested in future research. What can be concluded for the moment is that drawing valid inferences about teaching quality from questionnaire data is a challenge which still needs to be addressed through concerted effort.

ACKNOWLEDGEMENT

REFERENCES

Areepattamannil, S. (2012). Effects of Inquiry-Based Science Instruction on Science Achievement and Interest in Science: Evidence from Qatar. *The Journal of Educational Research*, *105*(2), 134–146. https://doi.org/10.1080/00220671.2010.533717

Asendorpf, J. B., & Conner, M. (2012). The g-factor of international cognitive ability comparisons: The homogeneity of results in PISA, TIMSS, PIRLS and IQ-tests across nations. *European Journal of Personality*, *119*(February), 108–119. https://doi.org/10.1002/per

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How People Learn: Brain, Mind, Experience, and School*. Washington DC: National Academies Press.

Cairns, D., & Areepattamannil, S. (2017). Exploring the Relations of Inquiry-Based Teaching to Science Achievement and Dispositions in 54 Countries. *Research in Science Education*, 1–23. https://doi.org/10.1007/s11165-017-9639-x

Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction*, *29*, 1–9. https://doi.org/10.1016/j.learninstruc.2013.07.001

Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2016). Experimental and Quasi-Experimental Studies of Inquiry-Based Science Teaching: A Meta-Analysis. *Review of Educational Research*, *82*(3), 300–329. https://doi.org/10.3102/0034654312457206

Gee, K. A., & Wong, K. K. (2012). A cross national examination of inquiry and its relationship to student performance in science: Evidence from the Program for International Student Assessment (PISA) 2006. *International Journal of Educational Research*, *53*, 303–318. https://doi.org/10.1016/j.ijer.2012.04.004

Jones, L. R., Wheeler, G., & Centurino, V. A. S. (2015). TIMSS 2015 Science Framework. In I. V. S. Mullis & M. O. Martin (Eds.), *TIMSS 2015 Assessment Frameworks* (pp. 29–59). Boston: Boston College, TIMSS & PIRLS International Study Center.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75–86. https://doi.org/10.1207/s15326985ep4102

Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras study: Investigating effects of teaching and learning in Swiss and German mathematics classrooms (pp. 137–160). Munster: Waxmann Verlag.

Lay, Y. F., Areepattamannil, S., Ng, K. T., & Khoo, C. H. (2015). Dispositions Towards Science and Science Achievement in TIMSS 2011: A Comparison of Eighth Graders in Hong Kong, Chinese Taipei, Japan, Korea, and Singapore. In *Science Education in East Asia* (pp. 591–605). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-16390-1_24

Lazonder, A. W., & Harmsen, R. (2016). Meta-Analysis of Inquiry-Based Learning: Effects of Guidance. *Review of Educational Research*, *86*(3), 681–718. https://doi.org/10.3102/0034654315627366

Liou, P.-Y., & Jessie Ho, H.-N. (2018). Relationships among instructional practices, students' motivational beliefs and science achievement in Taiwan using hierarchical linear modelling. *Research Papers in Education*, *33*(1), 73. https://doi.org/10.1080/02671522.2016.1236832

Lüdtke, O., Robitzsch, A., Trautwein, U., & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modeling. *Contemporary Educational Psychology*, *34*(2), 120–131. https://doi.org/10.1016/j.cedpsych.2008.12.001

Mayer, R. E. (2004). Should There Be a Three-Strikes Rule Against Pure The Case for Guided Methods of Instruction. *American Psychologist*, *59*(1), 14–19. https://doi.org/10.1037/0003-066X.59.1.14

Müller, K., Prenzel, M., Seidel, T., Schiepe-tiska, A., & Kjærnsli, M. (2016). Science Teaching and Learning in Schools: Theoretical and Empirical Foundations for Investigating Classroom-Level Processes. In S. Kuger, E. Klieme, N. Jude, & D. Kaplan (Eds.), *Assessing Contexts of Learning: An International Perspective* (pp. 423–446). Switzerland: Springer International. https://doi.org/10.1007/978-3-319-45357-6

Neumann, K., Kauertz, A., & Fischer, H. E. (2012). Quality of instruction in science education (pp. 1–4). https://doi.org/10.15713/ins.mmj.3

OECD. (2016). *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic and Financial Literacy*. Paris: OECD Publishing. https://doi.org/10.1787/9789264255425-en

Quintana, C., Zhang, M., & Krajcik, J. (2005). A framework for supporting metacognitive aspects of online inquiry through software-based scaffolding. *Educational Psychologist*, *40*(4), 235–244. https://doi.org/10.1207/s15326985ep4004_5

Sjøberg, S. (2018). The power and paradoxes of PISA: Should Inquiry-Based Science Education be sacrificed to climb on the rankings? *Nordic Studies in Science*, *14*(2), 186–202. https://doi.org/10.5617/nordina.6185

Stockard, J., Wood, T. W., Coughlin, C., & Rasplica Khoury, C. (2018). The Effectiveness of Direct Instruction Curricula: A Meta-Analysis of a Half Century of Research. *Review of Educational Research*, *XX*(X), 1–29. https://doi.org/10.3102/0034654317751919

Wilson, C. D., Taylor, J. A., Kowalski, S. M., & Carlson, J. (2010). The relative effects and equity of

inquiry-based and commonplace science teaching on students' knowledge, reasoning, and argumentation. *Journal of Research in Science Teaching*, *47*(3), 276–301. https://doi.org/10.1002/tea.20329

Wu, M. (2009). A critical comparison of the Contents of PISA and TIMSS Mathematics Assessments.

Zhang, L. (2016). Is Inquiry-Based Science Teaching Worth the Effort? *Science & Education*, *25*(7–8), 897–915. https://doi.org/10.1007/s11191-016-9856-0

IEA PROGRAM

Researching education, improving learning

8th IEA International Research Conference

Copenhagen, Denmark

Workshops: 24 - 25 June 2019
Conference: 26 - 28 June 2019

IRC 2019

IEA

DANISH SCHOOL OF EDUCATION
AARHUS UNIVERSITY

Dear participants,

We are sincerely glad to have the opportunity to welcome you to Denmark and the Danish School of Education—also known in Denmark, on our campus in Copenhagen, and beyond, as DPU. You are not only visiting the 8th IEA International Research Conference but also one of Denmark's largest and strongest university environments for research within the field of education and educational theory (pedagogy). With our 155 scholars and 90 PhD students, DPU constitutes one of the largest education research environments in Europe. Hosting the conference is in line with DPU's international foundation and outlook to build strong scientific perspectives on educational theory. DPU is characterized as a place for interdisciplinary research aimed at specific educational fields of inquiry, with daycare and schools playing the largest and most central roles. We therefore hope you have some fruitful days of knowledge sharing here at our campus, joining the 8th IEA International Research Conference.

Sincerely,

Christian Christrup Kjeldsen
Deputy Head of School (DPU) and Centre Director and Director for Research at the National Centre for School Research (NCS)

DANISH SCHOOL OF EDUCATION
AARHUS UNIVERSITY

CONTENTS

PROGRAM SCHEDULE

Monday, 24 June 2019

08:00 – 09:00	Workshop registration				
09:00 – 10:30	Workshop 1	Workshop 2	Workshop 3	Workshop 4	Workshop 5
10:30 – 10:45	Coffee / tea break				
10:45 – 12:15	Workshop 1	Workshop 2	Workshop 3	Workshop 4	Workshop 5
12:15 – 13:15	Lunch				
13:15 – 14:45	Workshop 1	Workshop 2	Workshop 3	Workshop 4	Workshop 5
14:45 – 15:00	Coffee / tea break				
15:00 – 16:30	Workshop 1	Workshop 2	Workshop 3	Workshop 4	Workshop 5

Tuesday, 25 June 2019

09:00 – 10:30	Workshop 1	Workshop 2	Workshop 3	Workshop 4	Workshop 5
10:30 – 10:45	Coffee / tea break				
10:45 – 12:15	Workshop 1	Workshop 2	Workshop 3	Workshop 4	Workshop 5
12:15 – 13:15	Lunch				
13:15 – 14:45	Workshop 1	Workshop 2	Workshop 3	Workshop 4	
14:45 – 15:00	Coffee / tea break				
15:00 – 16:30	Workshop 1	Workshop 2	Workshop 3	Workshop 4	

Wednesday, 26 June 2019

08:00 – 15:30	Conference registration			
09:00 – 09:30	Opening ceremony (Room: A220)			
09:30 – 11:00	Keynote 1			
11:00 – 11:30	Coffee / tea break			
11:30 – 12:30	Plenary panel			
12:30 – 13:30	Lunch			
13:30 – 15:00	Session 1A	Session 1B	Session 1C	Session 1D
15:00 – 15:30	Coffee / tea break			
15:30 – 17:00	Session 2A	Session 2B	Session 2C	Session 2D
18:00 – 19:30	Reception			

Thursday, 27 June 2019

08:30 – 15:30	Conference registration			
09:00 – 10:30	Keynote 2			
10:30 – 11:00	Coffee / tea break			
11:00 – 12:00	Plenary panel			
12:00 – 13:00	Lunch			
13:00 – 14:30	Session 3A	Session 3B	Session 3C	Session 3D
14:30 – 15:00	Coffee / tea break			
15:00 – 16:30	Session 4A	Session 4B	Session 4C	Session 4D

Friday, 28 June 2019

08:30 – 09:00	Conference registration			
09:00 – 10:30	Keynote 3			
10:30 – 11:00	Coffee / tea break			
11:00 – 12:30	Session 5A	Session 5B	Session 5C	Session 5D
12:30 – 13:30	Lunch			
13:30 – 15:00	Session 6A		Session 6C	Session 6D
15:00 – 15:30	Closing ceremony (Room: A220)			

PROGRAM OVERVIEW

Pre-Conference Workshops
Monday & Tuesday, 24-25 June 2019, 09:00 – 16:30*

Workshop 1	Using large-scale assessment data to inform policy and practice Instructor: *David Rutkowski and Sabine Meinck*	Room: A401 Abstract page: 14
Workshop 2	Multilevel modeling with IEA data Instructor: *Leslie Rutkowsk and Maria Magdalena Isac*	Room: A403 Abstract page: 14
Workshop 3	Assessment design, item response theory, and proficiency estimation Instructor: *Eugenio Gonzalez*	Room: A412 Abstract page: 15
Workshop 4	Response styles in large-scale assessment Instructor: *Jamis He Jia* (In memory of *Fons van de Vijver*)	Room: A414 Abstract page: 15
Workshop 5*	Analyzing data from international large-scale assessments using R Instructor: *Emmanuel Sikali, Paul Bailey, and Ting Zhang*	Room: A405 Abstract page: 15

*Workshop 5 ends at 12:15 on Tuesday 25 June.

Keynote presentations

Wednesday 26 June 2019 09:30 – 11:00	Keynote 1 Global attitudes and perceptions of social justice among youth: When no (in)differences make the difference Speaker: *Christian Christrup Kjeldsen* Chair: *Thierry Rocher*	Room: A220 Abstract page: 16
Thursday 27 June 2019 09:00 – 10:30	Keynote 2 How can IEA make a difference in measuring and monitoring learning in the 2030 agenda for sustainable development? Speaker: *Aaron Benavot* Chair: *Andrea Netten*	Room: A220 Abstract page: 16
Friday 28 June 2019 09:00 – 10:30	Keynote 3 For a fact-based worldview Speaker: *Anna Rosling Rönnlund* Chair: *Dirk Hastedt*	Room: A220 Abstract page: 17

Plenary sessions

Wednesday 26 June 2019 11:30 – 12:30	Panel on national challenges, regional perspectives, and global influences: The contribution of IEA studies to the better understanding of civic and citizenship education over the decades Panelists: *Gabriella Agrusti, Jens Bruun, Cristián Cox, Ines Elezović, and Wolfram Schulz* Moderator: *Ralph Carstens*	Room: A220 Abstract page: 17
Thursday 27 June 2019 11:00 – 12:00	Perspective on PIRLS and ePIRLS 2016 (Symposium) Chair: *Michael O Martin* Panelists: *Jan Mejding, Marc Colmant, Cecilia Stenman, and Bethany Fishbein* Discussant: *Ina Mullis*	Room: A220 Abstract page: 17

Closing ceremony

Friday 28 June 2019 15:00 – 15:30	Chair: Christian Christrup Kjeldsen Rapporteurs: Ina Mullis (Strand A: TIMSS and PIRLS), Monica Rosen (Strand B: Methodology), Andres Sandoval-Hernandez (Strand C: ICCS), and Jeppe Bundsgaard (Strand D: Working with IEA data)	
	The ceremony will close with an invitation to IEA IRC 2021	Room: A220

Session 1A: PIRLS and TIMSS
Teacher characteristics and instructional quality
Chair: *Trude Nilsen*
Discussant: *Martin Hooper*

Room: A401
Abstract page: 18

Examining reading instruction that works: Teacher profiles across 14 countries in PIRLS | *Joshua McGrane, Therese N. Hopfenbeck, Grace Grima, Kit Double, and Jenny Lenkeit*

Comparing teacher recruitment policies in primary education in Iran | *Masoud Kabiri and Ebrahim Talaei*

The meaning of instructional quality for reading competence in diverse classrooms | *Franziska Schwabe and Nele McElvany*

Session 1B: Methodology
New approaches to analyze IEA data
Chair: *Jan-Eric Gustafsson*

Room: A403
Abstract page: 19

Optimally predictive cross-country growth models with applications to TIMSS | *David Kaplan and Agnes Stancel-Piątak*

An examination of possible explanatory variables predicting the return of the TIMSS Early Learning Survey | *Nicole Wernert and Renee Sze Leung Kwong*

Predicting mathematics achievement: A machine learning approach using TIMSS 2015 US national public-use data | *Yuqi Liao and Trang Nguyen*

Multilevel modeling with large-scale assessment data: Uncovering sampling-related challenges | *Sabine Meinck*

Session 1C (Panel): ICCS
Exploring the role of ILSAs as change agents. Reflections on secondary data analysis of ICCS
Chair: *David Rutkowski*
Panelists: *Ralph Carstens, Christian Christrup Kjeldsen, Eva Klemenčič, and Maria Teresa Meléndez Irigoyei*

Room: A220
Abstract page: 20

Teaching tolerance in a globalized world | *Andres Sandoval-Hernandez, Maria Magdalena Isac, and Daniel Miranda*

Civic attitudes and behavioral intentions among 14-year-olds. How can education make a difference towards a more democratic and cohesive Europe? | *Zsuzsa Blasko, Patricia Dinis Mota da Costa, and Esperanza Vera-Toscano*

Session 1D: Engaging with IEA data
Panel on open source publishing with IEA
Moderator: *Dirk Hastedt*
Panelists: *Leslie Rutkowski, Seamus Hegarty, and Claudia Acuna*

Room: A405
Abstract page: 21

Session 2A: TIMSS and ICCS
Students' attitudes and achievement in TIMSS, TIMSS Advanced mathematics, and ICCS
Chair: *Michael O Martin*

Room: A220
Abstract page: 21

Relationships between students' attitudes and beliefs, and achievement in advanced mathematics | *Laura Palmerio and Elisa Caponera*

Meaningful clusters of eighth grade students in 2015 TIMSS mathematics using motivation variables | *Michalis P. Michaelides, Elena Papanastasiou, Gavin T.L. Brown, Hanna Eklöf, Militsa Georgieva Ivanova, and Anastasios Markitsis*

The role of parents' literacy attitudes on children's reading achievement (PIRLS 2016) | *Virginie Dupont, Dominique Lafontaine, Patricia Schillings, Stéphanie Géron, and Charlotte Dejaegher*

The relationship between student engagement and achievement across countries within regions using latent class analysis | *Renee Sze Leung Kwong and Greg Thomas Macaskill*

Session 2B (Symposium): Methodology
The future of international assessment? The promise and challenge of a multistage design in IEA studies
Chair: *Leslie Rutkowski*
Discussants: *Sabine Meinck and Eugenio Gonzalez*

Room: A403
Abstract page: 22

Parameter estimation stability and probabilistic routing in MST | *Leslie Rutkowski and David Rutkowski*

Module length and routing methods in MST | *Dubravka Svetina, Leslie Rutkowski, and David Rutkowski*

Differential item functioning in MST | *Montesrrat Valdivia, Leslie Rutkowski, and David Rutkowski*

On test taking motivation in MST | *Yuan-Ling Liaw, Leslie Rutkowski, and David Rutkowski*

Session 2C: ICCS
Civic attitudes in a changing world
Chair: *Wolfram Schulz*

Room: A405
Abstract page: 24

Attitudes towards minorities: Changes from 2009 to 2016 using data from IEA's ICCS | *Clara Beyer and Falk Brese*

Proximity to refugee centers and attitudes toward immigration | *Diego Cortes and Mojca Rozman*

Civic knowledge and civic self-efficacy: How do they contribute to the promotion of democratic attitudes? | *Zsuzsa Blasko and Patricia Dinis Mota da Costa*

Explaining variation in teacher-reported student activities in the community | *Falk Brese and Julian Fraillon*

Session 2D: Engaging with IEA data
Roundtable on national experiences and use of IEA studies for teacher professional development
Chair: *Josef Basl*
Discussants: *Sue Thomson, Khaled Temsah, Elena Papanastasiu, Dana Pražáková, and Franck Salles*

Room: A401
Abstract page: 25

Session 3A (Symposium): TIMSS
20 years of international large-scale assessments: How time series and trends can inform policymaking
Chair: *Rolf Strietholt*
Discussant: *Jan-Eric Gustafsson*

Room: A401
Abstract page: 26

Predicting change in mathematics achievement in Norway over time | *Rolf Vegar Olsen and Sigrid Blömeke*

Has equity changed in Norway over the last decades? | *Trude Nilsen, Julius Björnsson, and Rolf Vegar Olsen*

Twenty years of science motivation mirrored through TIMSS: Example of Norway | *Hege Kaarstein and Trude Nilsen*

Session 3B: Methodology
Methodological issues in large-scale assessments: Challenges and solutions (1)
Chair: *Julius Björnsson*

Room: A403
Abstract page: 27

Achievement gaps between immigrant and native students over time: Evidence from 20 years of TIMSS | *Ebru Erberber, Yifan Bai, Yuan Zhang, Frank Fonseca, Yemurai Tsokodayi, and Marissa Hall*

Aiming for measures of long-term trends in Grade 8 mathematics: Bridges from SIMS to TIMSS 2015 | *Monica Rosén, Erika Majoros, Stefan Johansson, and Jan-Eric Gustafsson*

Sensitivity of the RMSD for detecting item-level misfit in low-performing countries | *Jesper Tijmstra, Yuan-Ling Liaw, Maria Bolsinova, Leslie Rutkowski, and David Rutkowski*

A new powerful validation procedure to guarantee comparability in international attitude and test scales: 'Alignment' | *Ingrid Munck*

Session 3C: ICCS
Social media and civics culture: Evidence from ICCS 2016
Chair: *Julian Fraillon*
Discussant: *Maria Magdalena Isac*

Room: A405
Abstract page: 28

Social media harm institutional trust and thereby civic culture in Europe | *Mikael Sandberg*

Young people's use of social media and internet for civic engagement in 24 countries | *Wolfram Schulz and Tim Friedman*

School, church, and media: Explaining differences in civic knowledge in the bilingual school system of Estonia | *Triin Lauri, Kaire Põder, and Anu Toots*

Session 3D: Engaging with IEA data
A symposium on *IEA Compass: Briefs in Education* series
Chair: *David Rutkowski*
Discussant: *Dirk Hastedt*

Room: A220
Abstract page: 29

PIRLS for teachers: Making PIRLS results more useful for practitioners | *Therese Hopfenbeck and Jenny Lenkeit*

Preparing the ground: The importance of early learning activities at home for fourth grade student achievement | *Sabine Meinck and Agnes Stancel-Piqtak*

Is democracy overrated? Latin American students' support for dictatorships | *Andres Sandoval-Hernandez*

No one likes a bully: How systematic is international bullying in fourth grade? | *Leslie Rutkowski and David Rutkowski*

Session 4
Thursday, 27 June 2019, 15:00 – 16:30

Session 4A: TIMSS and PIRLS
Explaining the gaps: Factors of student achievement
Chair: *Falk Brese*

Room: A401
Abstract page: 30

Differences in students' and teachers' characteristics between high and low performing classes in Slovakia | *Samo Varsik*

Reading comprehension growth from PIRLS Grade 4 to Grade 6 | *Bieke De Fraine, Koen Aesaert, and Jan Van Damme*

Nordic students' achievement and school effectiveness in TIMSS 2015 | *Marie Wiberg and Ewa Rolfsman*

Birth month and mathematics performance relationships in Norway | *Annette Hessen Bjerke, Elisabeta Eriksen, Andre Rognes, and Bjørn Smestad*

Session 4B: Methodology
Methodological issues in large-scale assessments: Challenges and solutions (2)
Chair: *Monica Rosén*

Room: A403
Abstract page: 31

Measurement invariance of mathematics attitudes scale across the Gulf Cooperation Council countries | *Abeer A. Alamri*

Item parameter equivalence in TIMSS 2015 | *Huseyin Husnu Yildirim, Selda Yildirim, and Norman Verhelst*

Performance differences between TIMSS's paper-pencil test and tablet test of Dutch Grade 4 students | *Eva R. Hamhuis, Martina R.M. Meelissen, and Cees A.W. Glas*

How to use the world as a laboratory: Teachers' role in enhancing equity | *Trude Nilsen and Jan-Eric Gustafsson*

Session 4C: ICCS
Regional analysis of ICCS data
Chair: *Andres Sandoval-Hernandez*

Room: A220
Abstract page: 32

Change or stability? Civic knowledge, classroom climate, and expected political participation in Latin American schools | *Daniel Miranda and Juan Carlos Castillo*

Young people's support for authoritarianism, trust in institutions, and expected political participation in five Latin American countries | *Wolfram Schulz*

Latin American students' attitudes toward the use of violence | *Citlalli Sanchez-Alvarez*

What you know and what you believe: Explaining tolerance of corruption in Latin American students | *Diego Alonso Carrasco Ogaz, Robin Banerjee, Ernesto Treviño, and Cristóbal Villalobos*

Session 4D: Engaging with IEA data
Innovative approaches to analyzing IEA data
Moderator: *Sabine Meinck*

Room: A405
Abstract page: 33

Country-level trends in early literacy activities: Predicting gains in PIRLS reading? | *Martin Hooper*

Analyzing trends in socioeconomic inequality and educational outcomes: Results and methodological strategies | *Yifan Bai*

Does computer access equal skill or does practice make perfect? | *Yemurai Tsokodayi*

Analyzing ePIRLS results internationally after adjusting for students' offline reading ability | *Martin Hooper*

Analyzing IEA data in R: Introducing the EdSurvey package and its direct estimation function | *Paul Bailey*

Session 5A: TIMSS, PIRLS, and ICILS
Utilizing in-depth analysis of large-scale assessment data to improve teaching
Chair: *Rolf Strietholt*

Room: A401
Abstract page: 35

Clarifying TIMSS Advanced mathematics 2015 results: A didactical approach through levels of mathematical knowledge operation | *Franck Salles*

Differential item functioning as a pedagogical tool | *Jeppe Bundsgaard*

Nerds or polymaths? Performance profiles at the end of primary education | *Olesya Gladushyna and Rolf Strietholt*

I know I can, but do I have the time? The role of teachers' self-efficacy and perceived time constraints in implementing cognitive-activation strategies | *Nani Teig, Ronny Scherer, and Trude Nilsen*

Session 5B: ICILS
Schooling and learning in the Digital Age: Evidence from ICILS 2013
Chair: *Julian Fraillon*
Discussant: *Eva Klemenčič*

Room: A220
Abstract page: 36

ICT-related school culture and differences in students' CIL: Results from IEA-ICILS 2013 | *Birgit Eickelmann, Julia Gerick, and Jeppe Bundsgaard*

Explaining differences in students' computer and information literacy by analyzing response times in IEA-ICILS 2013 | *Corinna Massek, Kerstin Drossel, and Birgit Eickelmann*

Gender differences in computer and information literacy: Some findings from ICILS 2013 | *Sue Thomson, Eveline Gebhardt, Kylie Hillman, and John Ainley*

Session 5C: ICCS
Civics and citizenship education in countries around the world
Chair: *Jens Bruun*

Room: A405
Abstract page: 37

Cross-national priorities for civic and citizenship education | *Tim Friedman, Gabriella Agrusti, Valeria Damiani, and Jennifer Hong*

Apathetic, distrustful, interested, passive, institutional, and participatory: Profiles of youth political participation in Latin America | *Ernesto Treviño, Diego Alonso Carrasco Ogez, Cristóbal Villalobos, and Natalia López*

Aiming for sense of citizenship across the Nordic countries: Comparing priorities in citizenship education 2009-2016 | *Idunn Seland, Lihong Huang, Jens Bruun, Cecilia Arensmeier, and Jan Markus Löfström*

Italian teachers' preparation and training to teach civic and citizenship education topics and skills: Findings from ICCS 2016 | *Valeria Damiani, Elisa Caponera, and Bruno Losito*

Session 5D*
Poster session
Chair: *Seamus Hegarty*
Discussant: *Lauren Musu*

Room: A403
Abstract page: 38

Analysis of TIMSS 2015 framework: Input to the development of internationally-benchmarked 21st century Philippine assessment | *Marilyn Ubina Balagtas, Danda Crimelda Buhain Garcia, and Dexter Cheng Ngo*

An application of PRMSE to evaluate subscale score value in TIMSS 2015 fourth grade mathematics | *Kondwani Kajera Mughogho*

Modeling students' attitudes about learning mathematics in TIMSS 2015 context questionnaire | *Militsa Georgieva Ivanova and Michalis P. Michaelides*

Perceived math self-efficacy as predictors of pursuing a STEM degree | *Richard Carlos Limtuatco Velasco, Lee Kenneth Jones, and Yujiro Fujiwara*

Achievement profiles in mathematics and science among Flemish fourth graders: An exploratory analysis | *Lies Appels, Jerich Faddar, Sven De Maeyer, and Peter Van Petegem*

Norwegian science teachers' content coverage in middle school: Patterns and factors | *Stephan Daus*

Citizenship through the IEA waves: Cross-cohort and country measurement invariance of two citizenship dimensions | *Federico Viertel-Arze*

Measurement invariance in large-scale assessments using group-level fit: Patterns across scales and countries | *Janine Buchholz and Johan Braeken*

Two decades of ILSAs: Meta-interpretation of student achievement results via comparison matrix | *Mariusz Gałczyński*

* Each contributor to the session will give a short 3 minute presentation of their poster. One poster will be awarded the Constantinos Papanastasiou Poster Prize. For more information, please see the session abstract.

Session 6A: TIMSS
Teaching and instruction
Chair: *Franck Salles*

Room: A401
Abstract page: 42

Science teaching practices and learning outcomes in high-performing Asian countries: Comparing TIMSS and PISA 2015 | *Anindito Aditomo*

School autonomy and the distribution of teaching quality: An international comparative perspective | *Leah Natasha Glassow-Hill, Kajsa Yang Hansen, and Emilie Franck*

Instructional sensitivity of the TIMSS science test: Adjacent-grades within schools | *Stephan Daus, Agnes Stancel-Piątak, and Johan Braeken*

Differences in determinants for ICT use and students' achievement in mathematics: Secondary analyses of TIMSS 2015 | *Julia Gerick, Birgit Eickelmann, and Kerstin Drossel*

Session 6C
Socioeconomic background and student achievement: TIMSS and PIRLS
Chair: *Jan Mejding*
Discussant: *Plamen Mirazchiyski*

Room: A403
Abstract page: 43

Deconstruction of the negative social heritage? A search for variables confounding the simple relation between socioeconomic status and student achievement | *Rune Muller Kristensen*

Effects of early tracking on performance and inequalities in achievement: Combined evidence from PIRLS, TIMSS, and PISA | *Andrés Strello, Rolf Strietholt, Charlotte Siepmann, and Isa Steinmann*

High achievement in mathematics and science: A multilevel analysis of TIMSS 2015 data for Ireland | *Vasiliki Pitsia, Anastasios Karakolidis, and Gerry Shiel*

Session 6D: Engaging with IEA data
Free online data platforms and tools for education researchers
Authors: *Falk Brese, Yemurai Tsokodayi, and Yuqi Liao*

Room: A405
Abstract page: 44

LIST OF ATTENDING AUTHORS, CHAIRS, AND DISCUSSANTS

(c) = chair
(d) = discussant
(m) = moderator
(p) = panelist

Name	Institution/organization	Section
Acuna, Claudia	Springer	1D(p)
Aditomo, Anindito	Deutsches Institut für Internationale Pädagogische Forschung, Germany; University of Surabaya, Indonesia	6A
Aesaert, Koen	KU Leuven, Belgium	4A
Agrusti, Gabriella	LUMSA University, Italy	Plenary 1, 5C
Alamri, Abeer A.	National Center for Assessment, Qiyas, Saudi Arabia	4B
Appels, Lies	University of Antwerp, Belgium	5D
Arensmeier, Cecilia	Örebro University, Sweden	5C
Bai, Yifan	American Institutes for Research, US	3B, 4D
Bailey, Paul	American Institutes for Research, US	Workshop 5, 4D
Balagtas, Marilyn Ubina	Philippine Normal University, Philippines	5D
Basl, Josef	Czech School Inspectorate, Czech Republic	2D(c)
Benavot, Aaron	University of Albany, SUNY, US	Keynote 2
Björnsson, Julius	University of Oslo, Norway	3A, 3B(c)
Blasko, Zsuzsa	European Commission, Joint Research Centre, Italy	1C, 2C
Blömeke, Sigrid	University of Oslo, Norway	3A
Brese, Falk	IEA	2C, 4A(c), 6D
Bruno, Losito	Roma Tre University, Italy	5C
Bruun, Jens	Aarhus Universitet, Denmark	Plenary 1, 5C(c)
Buchholz, Janine	German Institute for International Educational Research, Germany	5D
Bundsgaard, Jeppe	Aarhus Universitet, Denmark	5A, 5B
Caponera, Elisa	INVALSI, National Institute for the Educational Evaluation of Instruction and Training, Italy	2A, 5C
Carstens, Ralph	IEA	Plenary 1(m), 1C(p)
Colmant, Marc	French Ministry of National Education and Youth, France	Plenary 2
Damiani, Valeria	LUMSA University, Italy	5C
Daus, Stephan	University of Oslo, Norway	5D, 6A
Dinis Mota da Costa, Patricia	European Commission, Joint Research Centre, Italy	1C, 2C
Dupont, Virginie	University of Liège, Belgium	2A
Eickelmann, Birgit	Paderborn University, Germany	5B, 6A
Eklöf, Hanna	Umeå University, Sweden	2A
Elezović, Ines	National Centre for External Evaluation of Education, Croatia	Plenary 1
Faddar, Jerich	University of Antwerp, Belgium	5D
Fishbein, Bethany	Boston College, US	Plenary 2
Fraillon, Julian	Australian Council for Educational Research, Australia	2C, 3C(c), 5B(c)
Franck, Emilie	IEA	6A
Friedman, Tim	Australian Council for Educational Research, Australia	3C, 5C
Gałczyński, Mariusz	McGill University, US	5D
Gerick, Julia	Universität Hamburg, Germany	5B, 6A
Gladushyna, Olesya	Technische Universität Dortmund, Germany	5A
Glassow-Hill, Leah Natasha	University of Gothenburg, Sweden	6A
Gonzalez, Eugenio	ETS and IERI (IEA-ETS Research Institute)	Workshop 3, 2B(d)
Grima, Grace	Pearson, UK	1A
Gustafsson, Jan-Eric	University of Gothenburg, Sweden	1B(c), 3A(d), 3B, 4B
Hansen, Kajsa Yang	University of Gothenburg, Sweden	6A
Hastedt, Dirk	IEA	Keynote 3(c), 1D(m), 3D(d)
He Jia, Jamis	Tilburg University, the Netherlands	Workshop 4
Hegarty, Seamus	Publication and Editorial Committee, IEA	1D(p), 5D(c)
Hooper, Martin	American Institutes for Research, US	1A(d), 4D
Hopfenbeck, Therese N.	University of Oxford, UK	1A, 3D
Huang, Lihong	OsloMet—Oslo Metropolitan University, Norway	5C
Isac, Maria Magdalena	University of Groningen, the Netherlands	Workshop 2, 1C, 3C(d)
Ivanova, Militsa Georgieva	University of Cyprus, Cyprus	2A, 5D
Johansson, Stefan	University of Gothenburg, Sweden	3B
Kaarstein, Hege	University of Oslo, Norway	3A
Kaplan, David	University of Wisconsin-Madison, US	1B

Karakolidis, Anastasios	Dublin City University, Ireland	6C
Kjeldsen, Christian Christrup	Aarhus Universitet, Denmark	Keynote 1, 1C(p)
Klemenčič, Eva	Educational Research Institute, Slovenia	1C(p), 5B(d)
Kristensen, Rune Muller	Danish School of Education, Denmark	6C
Kwong, Renee Sze Leung	Australian Council for Educational Research, Australia	1B, 2A
Lauri, Triin	Tallinn University, Estonia	3C
Liao, Yuqi	American Institutes for Research, US	1B, 6D
Liaw, Yuan-Ling	University of Oslo, Norway	2B, 3B
Losito, Bruno	Roma Tre University, Italy	5C
Macaskill, Greg Thomas	Australian Council for Educational Research, Australia	2A
Markitsis, Anastasios	University of Cyprus, Cyprus	2A
Martin, Michael O	Boston College, US	Plenary 2(c), 2A(c)
Massek, Corinna	Paderborn University, Germany	5B
McGrane, Joshua	University of Oxford, UK	1A
Meelissen, Martina R.M.	University of Twente, the Netherlands	4B
Meinck, Sabine	IEA	Workshop 1, 1B, 2B(d), 4D(d)
Mejding, Jan	Aarhus Universitet, Denmark	Plenary 2, 6A(c)
Meléndez Irigoyen, Maria Teresa	Instituto Nacional para la Evaluación de la Educación, Mexico	1C(p)
Michaelides, Michalis P.	University of Cyprus, Cyprus	2A, 5D
Miranda, Daniel	Pontificia Universidad Católica de Chile, Chile	1C, 4C
Mirazchiyski, Plamen	Educational Research Institute, Slovenia	6C(d)
Mughogho, Kondwani Kajera	University of Oslo, Norway	5D
Mullis, Ina	Boston College, US	Plenary 2(d)
Munck, Ingrid	Gothenburg University, Sweden	3B
Musu, Lauren	IEA	5D(d)
Netten, Andrea	IEA	Keynote 2(c)
Nilsen, Trude	University of Oslo, Norway	1A(c), 3A, 4B, 5A
Ogaz, Diego Alonso Carrasco	Pontificia Universidad Católica de Chile, Chile	4C, 5C
Olsen, Rolf Vegar	University of Oslo, Norway	3A
Palmerio, Laura	INVALSI, National Institute for the Educational Evaluation of Instruction and Training, Italy	2A
Papanastasiou, Elena	University of Nicosia, Cyprus	2A, 2D(d)
Pitsia, Vasiliki	Dublin City University, Ireland	6C
Pražáková, Dana	DEPP, Ministry of Education, France	2D(d)
Rocher, Thierry	IEA	Keynote 1(c)
Rognes, Andre	OsloMet-Oslo Metropolitan University, Norway	4A
Rolfsman, Ewa	Umeå University, Sweden	4A
Rosén, Monica	University of Gothenburg, Sweden	3B, 4B(c)
Rosling Rönnlund, Anna	Gapminder	Keynote 3
Rozman, Mojca	IEA	2C
Rutkowski, David	Indiana University, US	Workshop 1, 1C(c), 2B, 3B, 3D(c), 3D
Rutkowski, Leslie	Indiana University, US	Workshop 2, 1D(p), 2B(c), 2B, 3B, 3D
Salles, Franck	DEPP, Ministry of Education, France	2D(d), 5A, 6A(c)
Sanchez-Alvarez, Citlalli	Instituto Nacional para la Evaluación de la Educación, Mexico	4C
Sandberg, Mikael	Halmstad University, Sweden	3C
Sandoval-Hernandez, Andres	University of Bath, UK	1C, 3D, 4C(c)
Scherer, Ronny	University of Oslo, Norway	5A
Schulz, Wolfram	Australian Council for Educational Research, UK	Plenary 1, 1C(c), 3C, 4C
Schwabe, Franziska	Technische Universität Dortmund, Germany	1A
Seland, Idunn	OsloMet-Oslo Metropolitan University, Norway	5C
Shiel, Gerry	Educational Research Centre, Ireland	6C
Sikali, Emmanuel	Institute of Education Science, NCES, US	Workshop 5
Smestad, Bjørn	OsloMet-Oslo Metropolitan University, Norway	4A
Strello, Andrés	Technische Universität Dortmund, Germany	6C
Stenman, Cecilia	Swedish National Agency for Education, Sweden	Plenary 2
Strietholt, Rolf	Technische Universität Dortmund, Germany	3A(c), 5A(c), 6C
Svetina, Dubravka	Indiana University, US	2B
Talaei, Ebrahim	Research Institute for Education, Iran	1A
Teig, Nani	University of Oslo, Norway	5A
Temsah, Khaled	Ministry of Education, UAE	2D(d),
Thomson, Sue	Australian Council for Educational Research, Australia	2D(d), 5B
Treviño, Ernesto	Pontificia Universidad Católica de Chile, Chile	4C, 5C
Tsokodayi, Yemurai	American Institutes for Research, US	3B, 4D, 6D
Valdivia, Montesrrat	Indiana University, US	2B

Varsik, Samo	Ministry of Education, Science, Research, and Sport of the Slovak Republic, Slovak Republic	4A
Velasco, Richard Carlos Limtuatco	Texas Tech University, US	5D
Verhelst, Norman	Eurometrics	4B
Viertel-Arze, Federico	Universidad Santo Tomás, Chile	5D
Villalobos, Cristóbal	Pontificia Universidad Católica de Chile, Chile	4C, 5C
Wernert, Nicole	Australian Council for Educational Research, Australia	1B
Wiberg, Marie	Umeå University, Sweden	4A
Yildirim, Huseyin Husnu	Bolu Abant Izzet Baysal University, Turkey	4B
Yildirim, Selda	Bolu Abant Izzet Baysal University, Turkey	4B
Zhang, Ting	American Institutes for Research, US	Workshop 5

ABSTRACTS

Workshops

Workshop 1: Using large-scale assessment data to inform policy and practice
Instructors: David Rutkowski[1] and Sabine Meinck[2]
(Indiana University, US[1]; IEA[2])
Room: A401

Over the last 60 years, modern international large-scale assessments (ILSAs) have become influential educational policy tools, moving beyond their historical role as descriptive "snapshots" of educational systems. As both the numbers of participants involved and subjects assessed by ILSAs have increased, policymakers have become increasingly interested in how the data collected can help inform the policy. It is thus important to present policy-relevant information resulting from ILSAs in a brief and accessible format. This workshop will discuss how ILSA data can be used to inform policymakers and education practitioners. The instructors will provide a solid overview of how to construct a good policy brief, illustrated by practical examples. Participants should leave the workshop with a well-developed outline that they can later develop into a publishable document. The workshop will focus on the most recent cycles of the International Civic and Citizenship Education Study (ICCS) and the Progress in International Reading Literacy Study (PIRLS). This two-day workshop comprises a mix of both lectures and group work. Participants will have ample time for group discussion of possible topics for briefs, and to identify and develop research analyses of core relevance to their personal research interests. After a comprehensive introduction to the IEA IDB Analyzer, its application will be demonstrated using practical example analyses. Using simple tools that can handle the methodological issues automatically, participants will gain experience in working with IEA data and a better understanding of the possibilities and constraints for analyses. Participants will inspire each other by exchanging ideas on nationally-focused analysis, while the instructors will provide guidance and advice throughout on further possibilities or potential limitations. The workshop will showcase successful examples of how to analyze the data from IEA studies and communicate the results to policymakers and education practitioners.

Prerequisites:
Participants should possess a working knowledge of basic statistics, and will need to bring their own laptop PC with Microsoft Office and SPSS 16.0 or higher preinstalled.

Workshop 2: Multilevel modeling with IEA data
Instructors: Leslie Rutkowski[1] and Maria Magdalena Isac[2]
(Indiana University, US[1]; University of Groningen, the Netherlands[2])
Room: A403

Multilevel modeling (MLM) can be used to reflect the hierarchical structure of education systems and the clustered structure of the data. MLM usually provides more reliable and less biased results than traditional analysis methods for clustered data. This workshop is designed to introduce participants to the basic theory and application of multilevel modeling (MLM), focusing especially on those features that are particular to large-scale assessment data (such as weighting and scaling). Participants will gain basic practical experience of the application of two-level models to large-scale assessment data. The workshop begins with a methodological introduction to MLM and its underlying assumptions. Participants will receive practical training in how to use MPlus software for multilevel analysis of IEA data, and learn more about methods for model selection and hypothesis testing. The course considers methodological concepts related to the complex study and sampling design of large-scale assessments, and the course instructors will provide comprehensive advice on selecting the most appropriate multilevel modeling approach for data analysis.
The following topics will be covered:
- Methodological foundations of MLM;
- Two-level random coefficient models with L1 and L2 predictors;
- Short introduction to MPlus;
- Calculating the compositional effect;
- Centering approaches;
- Application of MLM to large-scale assessment data (incorporating weighting and plausible values); and
- Hypothesis testing and model selection.
The models will be presented, and workshop participants will then practice their implementation via a series of practical exercises using MPlus.

Prerequisites:
This workshop is aimed at individuals who already possess a working knowledge of large-scale assessment and a solid knowledge of intermediate statistics. Although no previous experience of MPlus is required, familiarity with syntax-based statistical software is beneficial. Participants must bring their own PC-compatible laptops with SPSS software (or similar alternative software that can be used for data preparation) preinstalled. A trial version on the MPlus software will be made available and used during the workshop. The workshop will be a mixture of lectures and hands-on training, to ensure participants gain both sound knowledge and practical expertise.

Workshop 3: Assessment design, item response theory, and proficiency estimation
Instructor: Eugenio Gonzalez
(ETS and IERI, ETS-IEA Research Institute)
Room: A412

This workshop provides an overview of the principles surrounding the design of large-scale assessments, the item response theory (IRT) models used to calibrate items, and the methodology used to assign proficiency estimates, also known as plausible values. Most of the workshop will be devoted to theoretical presentations, but there will be ample time allocated for discussion. At the end of the workshop, participants undertake practical example analyses with the software used in IEA studies. Topics covered during the workshop will include:

- Assessment design principles: examining both the advantages and disadvantages of current designs of large-scale assessments, and consequences for analysis;
- Principles of item response theory: including the advantages and disadvantages of different models, the limitations of IRT models, and using Parscale to compute item parameters and proficiency estimates; and
- Principles of population modeling and proficiency estimation: including review of procedures and techniques for carrying out conditioning, multiple imputations in large-scale assessments, plausible values and why they are useful, using DESI (Direct Estimation Software Interactive) to compute proficiency estimates and plausible values.

Prerequisites:
This workshop is aimed at individuals who already possess a working knowledge of large-scale assessment and some understanding of statistics. Participants must bring their own PC-compatible laptops with SAS or SPSS software preinstalled. Trial versions of software used for the practical examples will be provided during the workshop.

Workshop 4: Response styles in large-scale assessment
Instructor: Jamis He Jia (In memory of Fons van de Vijver)
(Tilburg University, the Netherlands)
Room: A414

This workshop addresses response styles in large-scale assessment. Attitudinal and behavioral self-reports, the common item types, are known to be influenced by response styles, such as acquiescence, midpoint responding, extremity responding, and social desirability. The workshop will start with an introduction in the topic, addressing the following issues:

- Response styles in historical perspective: styles as nuisance factors, models of social desirability;
- Direct and indirect assessments of response styles;
- Correlates of response styles at individual level;
- How many response styles are there? Towards an integrative view;
- The cross-cultural perspective on response styles: distortion or communication style?;
- Correlates of response styles at country level;
- Procedures for reducing response styles (such as Likert versus frequency scales, response standardization, anchoring vignettes, and statistical response style corrections); and
- Influence of corrections on factors structures and individual/cultural differences.

After this introduction, participants will be shown how to use an IEA data set to interactively address the topics, and will learn the answers to such questions as:

- How can we compute response style indices? Which response styles can typically be computed in IEA data sets?
- How large are individual and country differences in these styles?
- Are the response styles correlated at individual and country level? Can these be integrated?
- What are the correlates at individual level?

Internet sources of country data will be referenced, enabling correlates at country level to be computed.

Prerequisites:
SPSS and Mplus software will be used (other packages with similar functionality can also be used). It is encouraged that participants bring their own laptop that can be used to conduct all computations in the second part of the workshop.

Workshop 5: Analyzing data from international large-scale assessments using R
Instructors: Emmanuel Sikali[1], Paul Bailey[2], and Ting Zhang[2]
(Institute of Education Science, NCES, US[1]; American Institutes for Research, US[2])
Room: A405

Data from international large-scale assessment programs such as TIMSS are valuable resources for researchers studying issues related to education, and such data may be used to inform educational policy. Due to the scope, complexity and special features of large-scale assessment data, researchers often have to use several different software tools to access, manipulate, and analyze the data. Mindful of this, the US National Center for Education Statistics (NCES) has developed an R package called EdSurvey, designed specifically for the analysis of national and international large-scale assessment data. The primary purpose of this workshop is to introduce participants to the currently available functionality of EdSurvey and explain how this free-to-download R package may be used to analyze TIMSS and TIMSS-like data. This 1.5-day course will comprise both lectures and practical exercises. The instructors will begin by introducing participants to the unique design features of international large-scale assessment data, providing guidance on the data analysis strategies that researchers can employ, including the selection and use of appropriate plausible values, sampling weights, and variance estimation procedures (such

as the jackknife approach). Following this introduction, participants will undertake interactive, practical training in analyzing public-use TIMSS data files using the R package EdSurvey. The theoretical knowledge and analytic approach learned during this course may be applied to analyze other international large-scale assessment data with plausible values, such as TIMSS Advanced, PIRLS, ICILS, ICCS, and PISA data.

Prerequisites:
Participants should have basic knowledge of statistical techniques, including statistical inference and multiple regression analysis. It is helpful, though not requisite, to have some working knowledge of item response theory and sampling theory, as well as some familiarity with R software (e.g., participants may have already completed an entry-level training on R programming, or used R for basic programming, such as simple data cleaning or running a regression). As this workshop includes extensive practical content, participants must bring their own laptops with the latest version of R and R studio preinstalled.

Keynotes

Keynote 1: Global attitudes and perceptions of social justice among youth: When no (in)differences make the difference
Speaker: Christian Christrup Kjeldsen (Aarhus Universitet, Denmark)
Chair: Thierry Rocher (IEA)
Room: A220

By means of a secondary analysis of the IEA International Civic and Citizenship Education Study (ICCS) 2016, this keynote presentation will offer a change in analytical and statistical perspective. Instead of following the quest for statistically significant mean differences, it seeks to highlight the many similarities in attitudes and perceptions of issues related to social justice and diversity among youth—despite quite different national state cultural contexts and political landscapes. This keynote is theoretically guided by a Marxist perspective and the moral philosophical insights of Amartya Sen's Capability Approach, and empirically informed by analysis of data from ICCS. It endeavors to question whether the similarities we see in ICCS results, indicating minimal differences among youth attitudes and perceptions in regards to social justice, should in fact be interpreted as a slight hope for the future global society in times of crisis. ICCS categorizes students on a civic knowledge scale. For students to be recognized as meeting the basic threshold of proficiency levels they have to prove themselves capable of "[recognizing] the motivations of people engaged in activities that contribute to the common good." Interestingly this is well in line with the recognition of motivation found in the graduation paper of a young man, though several generations ago. The young graduate, Marx, argues: "[h]istory calls those men the greatest who have ennobled themselves by working for the common good ... If we have chosen the position in life in which we can most of all work for mankind, no burdens can bow us down, because they are sacrifices for the benefit of all."

Keynote 2: How can the IEA make a difference in measuring and monitoring learning in the 2030 Agenda for Sustainable Development?
Speaker: Aaron Benavot (University of Albany, SUNY, US)
Chair: Andrea Netten (IEA)
Room: A220

For more than five decades, the IEA has been at the forefront of understanding the processes by which, and the policy contexts in which, students gain mastery of important knowledge and relevant skills. In my view, several essential principles have undergirded the many comparative studies of academic achievement conducted by IEA researchers over the years. These include, for example, the value of enabling national education experts and curriculum specialists to discuss the advantages of different research designs; the importance of gathering systematic information from multiple sources (principals, teachers, and students) to better understand the influence of different conditions and factors on learning; the need to illuminate the links between policy intentions, school implementation, and what student know and can demonstrate proficiency in; and the importance of a broad view of student learning encompassing different domains and skills sets. Recently, in 2015, 193 governments agreed on a new development agenda, the 2030 Agenda for Sustainable Development, which includes a global goal in education (SDG 4) and 11 key targets, many of which focus on learning among children, youth, and adults. Representatives of the international education community not only adopted this ambitious agenda in education, but provided guidance and concrete ideas for implementing SDG 4 and its targets through the Education 2030 Framework for Action. They also established committees and platforms, which were tasked with devising new measurement and monitoring strategies of learning in multiple domains among school age children and adolescents. The challenges and stakes accompanying these international processes are quite significant. My presentation will consider ways that the IEA can inform and support these measurement and monitoring challenges, by drawing on its core principles, as well as its long history of assessing quality schooling and student learning. I intend to highlight specific ways that IEA researchers can contribute to the evolution of policy-oriented data and indicators and is relevant for national and regional policy leaders in education who seek to improve educational opportunity and outcomes, especially among poor, marginal, and vulnerable populations.

Keynote 3: For a fact-based worldview
Speaker: Anna Rosling Rönnlund (Gapminder)
Chair: Dirk Hastedt (IEA)
Room: A220

We have tested the public on the most basic global development trends and people seem to be completely misinformed about what the world looks like. Anna will dismantle the three mega-misconceptions that shape our worldview, showing what the world really looks like in a way that everyone can understand. The audience will learn how these results relate to and are relevant for IEA data.

Plenary sessions

Panel on national challenges, regional perspectives, and global influences: The contribution of IEA studies to the better understanding of civic and citizenship education over the decades
Panelists:
   Gabriella Agrusti (ICCS Associated Study Director; LUMSA University, Italy)
   Jens Bruun (National Research Coordinator; Aarhus Universitet, Denmark)
   Cristián Cox (Project Advisory Committee member; University Diego Portales, Chile)
   Ines Elezović (National Research Coordinator; National Centre for External Evaluation of Education, Croatia)
   Wolfram Schulz (International Study Director; Australian Council for Educational Research, UK)
Moderator: Ralph Carstens (IEA)
Room: A220

The IEA ICCS studies in civic and citizenship education pursue "moving targets"—across the years, in changing educational contexts, from local to global levels. They focus on planned and implemented learning opportunities to prepare students as active and informed citizens along with contextual factors assumed to be related to educational outcomes with the use of tailored approaches, such as regional questionnaires. Emphasis in these studies is on: i) knowledge requirements from local to global contexts, ii) competencies needed from passive to active citizenship, and iii) the role of schools from transmitting knowledge to guiding student activities. Measuring such multifaceted knowledge-attitude constructs presents challenges for educational researchers in the field of international large-scale assessments. Broader, socio-political contexts change and require adjustments. Major challenges related to civic and citizenship education lie in the global legitimation crisis of representative democracy, peaceful coexistence, social media, and environmental deterioration. Besides foundational knowledge, related views and behaviors drive the development of research in this area, as well as the need to report "differences that make the difference." This plenary panel brings together voices from Europe (Denmark, Croatia, and Italy), Latin America, the international study centers, and the project advisory committee, to discuss continuing and emerging issues from IEA CIVED 1999 to ICCS 2022. How has ICCS contributed to our understanding, debates, and policy reactions? How did, or could, school education react? How could ICCS 2022 contribute to finding better answers of national, regional, and global relevance, especially in the context of SDG 4.7?

Perspective on PIRLS and ePIRLS 2016 (Symposium)
Chair: Michael O Martin (Boston College, US)
Panelists:
   Jan Mejding (Aarhus Universitet, Denmark)
   Marc Colmant (French Ministry of National Education and Youth, France)
   Cecilia Stenman (Swedish National Agency for Education, Sweden)
   Bethany Fishbein (TIMSS & PIRLS International Study Center, Boston College, US)
Discussant: Ina Mullis (Boston College, US)
Room: A220

This session will focus on the main PIRLS 2016 results in Sweden, France, and Denmark, including how they were discussed and received in each of the countries. Sweden and Denmark had relatively high achievement in PIRLS 2016, with France performing in the middle ranges. Sweden and France have trend results over four cycles of PIRLS since 2001, both having a pattern of some declines. Sweden did show recent improvement between 2011 and 2016, especially at the upper benchmarks. Denmark has trend results over three cycles of PIRLS since 2006, showing few changes overall or at the benchmarks. Interestingly, Denmark administered PIRLS 2016 at the fourth grade together with PIRLS literacy 2016 at the third grade, which provided considerable information about gains in reading achievement between Grade 3 and Grade 4. Denmark and Sweden also participated in the inaugural assessment of ePIRLS 2016. ePIRLS is an innovative assessment of online informational reading in a simulated internet environment. Given the increasing emphasis on the ability to find, interpret, and use digital information, ePIRLS provided an interesting look into students' achievement in reading in a digital environment. When comparing achievement in informational reading, some countries had higher achievement in ePIRLS and some had higher achievement in PIRLS. Denmark and Sweden had very similar ePIRLS achievement. Denmark performed significantly better on ePIRLS informational reading than on the paper and pencil format. Sweden also showed an advantage on ePIRLS compared to PIRLS, but the difference was smaller than in Denmark. Finally, as a step toward fully integrating ePIRLS into the PIRLS assessment, and to prepare for reporting the results of PIRLS and digitalPIRLS in 2021, the TIMSS & PIRLS International Study Center investigated the relationship between the paper-based PIRLS and online ePIRLS informational reading constructs as well as the effect that ePIRLS would have on the PIRLS 2016 item response theory

achievement results. The results of a multidimensional item factor analysis provides evidence that the PIRLS and ePIRLS 2016 items together can be considered to measure a unidimensional reading construct, while achievement differences between PIRLS informational reading on paper and ePIRLS informational reading on the internet can be plausibly explained in terms of an administration mode effect.

Parallel sessions

Session 1A: PIRLS and TIMSS
Teacher characteristics and instructional quality
Chair: Trude Nilsen (University of Oslo, Norway)
Discussant: Martin Hooper (American Institutes for Research, US)
Room: A401

Examining reading instruction that works: Teacher profiles across 14 countries in PIRLS
Joshua McGrane[1], Therese N. Hopfenbeck[1], Grace Grima[2], Kit Double[1], and Jenny Lenkeit[1]
(University of Oxford, UK[1]; Pearson, UK[2])

Research on international large-scale assessments has neglected to investigate the role of differences at the classroom level in the variance in achievement between education systems. The classroom level is, however, of interest for policymakers, as it constitutes malleable structures and characteristics, which are closest to the individual level of achievement. This research uses data from PIRLS 2016 to investigate how information on instructional practices can be mapped into teacher profiles or typologies of teaching, and how these typologies are related to student achievement and motivational characteristics. We draw on theoretical frameworks that identify instructional engagement, student-centered learning, individual support, classroom management, and climate as characteristics of teaching quality. Instructional practices are often examined using individual variables and/or scales, assuming an "either-or" or "the-more-the-better" approach to teaching. This research applies Latent Profile Analysis to describe typologies of teaching that better reflect the fact that teachers apply a combination of practices in their classroom. Further regression analyses are conducted to evaluate the relationship between teaching typology/teacher profiles, motivational characteristics and reading achievement across 14 education systems participating in PIRLS 2016. Findings will enhance our understanding of differences in achievement between education systems and will provide an impetus for enhancing teacher qualifications at national levels.

Comparing teacher recruitment policies in primary education in Iran
Masoud Kabiri and Ebrahim Talaei
(Research Institute for Education, Iran)

The focus of this article is the comparison of diverse policies of teacher recruitment in primary education in Iran. Teachers were divided into four groups, according to their response to the national items regarding how they were recruited by the Ministry of Education. Then, their students' outcomes (science achievement and attitudes toward science) and their activities (confidence in teaching science and emphasis on investigation in the science classroom) were considered as dependent variables. Moreover, categorical regression was applied to analyze TIMSS 2015 data and statistical controlling was done by year of teaching and type of school. Results revealed that science achievement and attitude of students did not differ based on how their teachers were recruited. Additionally, there is little difference in teaching confidence between teachers after statistical controlling. Furthermore, teachers from Teacher Training College reported less emphasis on investigation in the classroom than Literacy Movement and Hourly Paid teachers. The findings question the appropriateness of teacher preparation. Improved quality of teacher education is recommended.

The meaning of instructional quality for reading competence in diverse classrooms
Franziska Schwabe and Nele McElvany
(Technische Universität Dortmund, Germany)

Reading is a focal competence that is acquired mostly in primary school. The heterogeneity in primary school classrooms makes it necessary to examine which aspects are particularly relevant for reading competence in this context. With respect to language classes, a very relevant aspect of classroom composition is the student's knowledge of the language of teaching. High proportions of students with a language minority background have been found to be detrimental to individual as well as classroom reading competence. High instructional quality in reading classes might be of great relevance especially in this context. Against this background the present study investigated the relationship between two key elements of instructional quality: (a) teaching reading strategies and (b) individualizing reading instruction, language minority classroom composition, and reading competence in Grade 4. The first research aim was to clarify the direct relation between aspects of instructional quality and students' reading competence on the classroom level. Secondly, we investigated an interaction effect between language minority composition and aspects of instructional quality on students' reading competence on the classroom level. Analyses were based on PIRLS 2006 data (German sample; N=7,767 students). Identified by asking the students about the amount of time they speak the language of the school at home with their family, 29.1 percent of the students had a language minority background. Moreover N=387 teachers were included in the study and reported on their instruction. Two two-level structural equation models (SEMs) were conducted using five imputed datasets with students' reading competence as a dependent variable. While a positive relation was found between amount of teaching reading strategies and students' reading competence on the classroom level, analyses revealed no significant relationship between individualization and

students' reading competence on the classroom level. The second SEM revealed a statistically significant interaction effect between amount of teaching reading strategies and proportion of language minority students on students' reading competence on the classroom level. Thereby, the assumption of differential meaning of instructional quality for diverse classrooms was supported. Results indicated that instructional quality related positively to students' reading achievement and might be a starting point to overcome achievement gaps between students with different language backgrounds.

Session 1B: Methodology
New approaches to analyze IEA data
Chair: Jan-Eric Gustafsson (University of Gothenburg, Sweden)
Room: A403

Optimally predictive cross-country growth models with applications to TIMSS
David Kaplan[1] and Agnes Stancel-Piątak[2]
(University of Wisconsin-Madison, US[1]; IEA[2])

Of critical importance to education policy is the monitoring of trends in educational outcomes over time. Indeed, the United Nations Sustainable Development Goals identified Goal 4 as focusing on quality education for all. For example, Goal 4.6 focuses specifically on achieving literacy and numeracy for men and women. With six cycles of TIMSS, these data enable cross-time comparisons at the country level. A careful analysis of trend data at the country level can allow policy-makers to assess cross-country progress and forecasts toward policy-relevant goals. However, to optimize the predictive capacity of forecasting models, it is statistically desirable to average over a large space of plausible forecasting models using the method of Bayesian model averaging. The focus of this paper is to present new developments and results on optimal cross-country growth regression using Bayesian model averaging, and the opportunities and challenges of using cross-country growth regressions of TIMSS data to construct forecasting models of policy-relevant educational outcomes. At present, we have fit the cross-country growth curve model to gender differences in 8th grade math to 21 countries with sufficient time-points. Our results indicate that a non-linear growth curve model best fits the observed math scores over six cycles of TIMSS. The model does a good job of forecasting the 2015 results based on previous findings, and ex-post forecast statistics and the Kullback-Liebler divergence measure indicate that the non-linear model fits the observed time trend quite well. Initial results comparing our approach to optimal prediction using Bayesian model averaging (BMA) versus ignoring model uncertainty, indicates a small advantage of BMA. Our findings to date indicate that TIMSS can be used to estimate cross-country forecasting models. Our results also suggest that accounting for forecasting model uncertainty provides small but noticeable benefits as measured by forecasting scoring rules. With well-calibrated growth models, researchers can, in principle, identify important and policy manipulable predictors of growth in educational outcomes. Our point of view is that the current design of international large-scale educational assessments, and TIMSS in particular, can be exploited to address important questions of international policy relevance, but to do so requires adopting a forecasting and prediction point of view within a Bayesian uncertainty quantification framework.

An examination of possible explanatory variables predicting the return of the TIMSS Early Learning Survey
Nicole Wernert and Renee Sze Leung Kwong
(Australian Council for Educational Research, Australia)

A key issue for researchers looking at parental involvement in education is getting access to parents and ensuring a respectable and non-biased response set. The Trends in International Mathematics and Science Study (TIMSS) includes a parent survey (also known as the Early Learning Survey) which includes some key variables of interest to researchers looking at parental involvement. However, several countries did not achieve adequate response rates, with some (including Australia and the Netherlands) having a response rate so low that the results were not reported. This paper looks at whether the response rate varies according to a number of variables that could potentially have explanatory value, including student achievement, student background variables, and various questionnaire variables. Data from six countries with low response rates for the Early Learning Survey (ELS) were used. It was found that there were differences in the response rate of the ELS across the categories of the variables of interest. In addition, mean achievement was higher for students whose parents had responded to the ELS than for those whose parents had not, and this relationship holds within the categories of the variables of interest. However, a logistic regression analysis found that the variables of interest did not improve the prediction of whether or not a parent returns the ELS. This indicates that demographic categories, or even student disengagement, were not the greatest influence on whether or not a parent survey was returned or not. This finding suggests that suggestions for increasing parent survey response rates that focus on demographic variables may only have limited effect on response rates.

Predicting mathematics achievement: A machine learning approach using TIMSS 2015 US national public-use data
Yuqi Liao and Trang Nguyen
(American Institutes for Research, US)

Using TIMSS 2015 US national public-use data, this study shows how machine learning could be applied to large-scale international studies. The analysis provides a comparison of six commonly used machine learning models in identifying students with low mathematics performance, which would allow educators to effectively allocate resources to intervene and thus improve their performance. The results show that logistic regression, elastic net, and extreme gradient boosting all perform well in terms of balanced accuracy and sensitivity. Leveraging the high dimensionality of the data, this research analyzes 142 student and school variables, and identifies characteristics that are associated with mathematics performance,

including students' home education resources, attitudes toward mathematics, and use and possession of computers or tablets.

Multilevel modeling with large-scale assessment data: Uncovering sampling-related challenges
Sabine Meinck
(IEA)

This paper reviews arising challenges when data from international large-scale assessments (ILSA) is used for multilevel analysis. Commonalities and differences of the sampling designs of ILSA that affect multilevel models are introduced. The article then determines how sampling designs actually mimic the hierarchies in the populations, and whether or not, or in which circumstances, these designs allow reasonable multilevel modeling of the population hierarchies. Special attention is given to the challenges of unequal selection probabilities and related sampling weights, and how these are currently included in multilevel models. The approaches of using total weights, split-weights for different hierarchical levels, and standardized weights will be discussed, emphasizing the differences in the proposed methodologies and current debates on this topic. Further, issues arising from the fact that clustering can occur at more than one sampling stage are addressed. In such designs, determining the different hierarchical levels is a concern. The paper reviews possible solutions, their advantages, and shortcomings. Moreover, attention is given to other topics resulting from the various sampling designs, such as handling of teacher and aggregate data. The contribution highlights the need of further research for informing current debates, and substantiates the plea for providing accessible guidelines addressed to researchers conducting multilevel analysis with ILSA data.

Session 1C (Panel): ICCS
Exploring the role of ILSAs as change agents. Reflections on secondary data analysis of ICCS
Chair: David Rutkowski (Indiana University, US)
Panelists: Ralph Carstens[1], Christian Christrup Kjeldsen[2], Eva Klemenčič[3], and María Teresa Meléndez Irigoyen[4]
(IEA[1]; Aarhus Universitet, Denmark[2]; Educational Research Institute, Slovenia[3]; Instituto Nacional para la Evaluación de la Educación, Mexico[4])
Room: A220

The aim of this session is to foster a productive and systematic conversation involving several stakeholders about the utility of secondary data analysis of international large-scale assessment (ILSA) data for informing education policy. We start with the assumption that ILSAs can play an important role in informing education policy and that they can be particularly useful for providing input in two stages of the education policy cycle: evaluation and agenda setting. For example, ILSAs can provide information regarding the implementation of policy goals but also highlight new areas that require attention and policy action. Nevertheless, translating scientific findings into policy implications is not always an easy task and such practices are not widespread. Our aim is therefore to systematically asses the feasibility of translating research results into policy implications. To this end, we focus on the International Civic and Citizenship Education Study (ICCS 2009 and 2016) and two science and policy reports focused on these data: 1) the *IEA Research for Education* report, 'Teaching tolerance in a globalized world' and 2) the Joint Research Commission, European Commission (EC), *Science for Policy* report, 'Civic attitudes and behavioral intentions among 14-year-olds. How can education make a difference towards a more democratic and cohesive Europe?' Both publications were commissioned and supported by stakeholders with an interest in science for policy: IEA and the EC. They both share an intended connection with important priority education policy goals at national and international levels (for example, the assessment of citizenship competences in youth and the evaluation of citizenship education programs) and include reflections on policy implications. To further assess the potential of these results, we aim to foster a systematic discussion of the findings involving the perspectives of several stakeholders. As much as possible, this session will therefore gather perspectives from: the funding agencies, the authors/researchers of the publications, the perspective of education policy analysts, and the perspective of national and European representatives/policy makers. All participants (presenters, discussant, panelist, and the audience) will reflect on a number of overarching objectives:
- Identifying the priority area(s) in education policy that motivated the work;
- Reflecting on the impact that this agenda had for the process of empirical data analysis;
- Mapping the main findings and reflecting on their potential utility (either for evaluative purposes and/or for agenda setting); and
- Identifying gaps in knowledge and topics/areas where further research input is needed.

Teaching tolerance in a globalized world
Andres Sandoval-Hernandez[1], Maria Magdalena Isac[2], and Daniel Miranda[3]
(University of Bath, UK[1]; University of Groningen, the Netherlands[2]; Pontificia Universidad Católica de Chile, Chile[3])

This open access thematic report identifies factors and conditions that can help schools and education systems promote tolerance in a globalized world. The IEA's International Civic and Citizenship Study (ICCS) is a comparative research program designed to investigate the ways in which young people are prepared to undertake their roles as citizens, and provides a wealth of data permitting not only comparison between countries but also comparisons between schools within countries, and students within countries. Advanced analytical methods provide insights into relationships between students' attitudes towards cultural diversity and the characteristics of the students themselves, their families, their teachers, and school principals. The rich diversity of educational and cultural contexts in the 38 countries who participated in ICCS 2009 are also acknowledged and addressed. Readers interested in civic education and adolescents' attitudes towards cultural diversity

will find the theoretical perspectives explored engaging. For readers interested in methodology, the advanced analytical methods employed present textbook examples of how to address cross-cultural comparability of measurement instruments and multilevel data structures in international large-scale assessments (ILSA). Meanwhile, those interested in educational policy should find the identification and comparison of malleable factors across education systems that contribute to positive student attitudes towards cultural diversity a useful and thought-provoking resource.

Civic attitudes and behavioral intentions among 14-year-olds. How can education make a difference towards a more democratic and cohesive Europe?
Zsuzsa Blasko, Patricia Dinis Mota da Costa, and Esperanza Vera-Toscano
(European Commission, Joint Research Centre, Italy)

Using data from the 2016 International Civic and Citizenship Education Study (ICCS), this policy report provides a detailed analysis of adolescents' civic attitudes and behavioral intentions, and the mechanisms shaping them, with a particular emphasis on the broader role of education. It provides an opportunity to obtain greater insights into the role that educational policy in general, and educational institutions in particular, may play in shaping civic outcomes. The report distinguishes between formal learning of civic issues, school democracy, and community work, and assesses their individual contributions to a range of non-cognitive civic outcomes of students across 13 EU Member States. It highlights notable variations across countries as well as across civic outcomes in these mechanisms. Still, common patterns are also identified, such as the importance of open classroom climate, the need for improving students' civic self-efficacy and the potential usefulness of community work.

Session 1D: Engaging with IEA data
Panel on open source publishing with IEA
Chair: Dirk Hastedt (IEA)
Presenters: Leslie Rutkowski[1], Seamus Hegarty[2], and Claudia Acuna[3]
(Indiana University, US[1]; Publication and Editorial Committee, IEA[2]; Springer[3])
Room: A405

This session will introduce the open source publishing arena, with a particular focus on IEA's partnership with Springer. We will focus on two main opportunities within IEA for open access publishing with Springer: submitting single papers (*Large-scale Assessments in Education* journal) and submitting proposals to author thematic reports on topics of wide interest and importance (*IEA Research for Education* series).

Session 2A: TIMSS and ICCS
Students' attitudes and achievement in TIMSS, TIMSS Advanced mathematics, and ICCS
Chair: Michael O Martin (Boston College, US)
Room: A220

Relationship between students' attitudes and beliefs, and achievement in advanced mathematics
Laura Palmerio and Elisa Caponera
(INVALSI, National Institute for the Educational Evaluation of Instruction and Training, Italy)

This study investigated the relationship between student self-beliefs and mathematics performance using a structural equation modeling approach. Italian students participating in TIMSS Advanced 2015 (N=1982; mean age, 18 and 10 months, ± 8) were considered. Students answered the TIMSS Advanced mathematics test and the international questionnaire that included questions regarding students' socioeconomic and cultural background, attitudes and beliefs (how they enjoy and value mathematics), and future expectation regarding their educational career. Furthermore, the same students participated in a national survey and answered questions regarding mathematics self-efficacy and anxiety. The results showed that all measures were significantly associated with the mathematics achievement score. Self-efficacy was the best predictor of mathematics performance. This study indicates that the beliefs system self-efficacy is based on, plays a key role in mathematics achievement. A fundamental aspect of beliefs is that they may be modified. Therefore, it is important to consider specific interventions in this direction in the teaching and learning process.

Meaningful clusters of eighth grade students in 2015 TIMSS mathematics using motivation variables
Michalis P. Michaelides[1], Elena Papanastasiou[2], Gavin T.L. Brown[3,4], Hanna Eklöf[4], Militsa Georgieva Ivanova[1], and Anastasios Markitsis[1]
(University of Cyprus, Cyprus[1]; University of Nicosia, Cyprus[2]; University of Auckland, New Zealand[3]; Umeå University, Sweden[4])

Items measuring motivation and affect are regularly administered in the TIMSS background questionnaires. These variables have been shown to be positively associated with achievement in mathematics and science. However, in models combining multiple predictors, not all of them are equally good predictors. Confidence in mathematics for example is a stronger predictor than enjoyment or value for the subject. In this study, a person-centered analytic approach was adopted: confidence, enjoyment, and value for mathematics were used as input variables in cluster analysis using data from TIMSS 2015 across 12 countries or jurisdictions. Results indicated that some clusters consisted of students who scored consistently high, moderate, or low on all three motivational variables. However, there were clusters with inconsistent ratings (for

example, students endorsing high value but low confidence and enjoyment). Clusters were compared with respect to gender composition, home educational resources, and average mathematics achievement. Systematic patterns appeared across the datasets and suggest that (a) achievement is associated with motivation, (b) not all motivation variables are equally important in predicting high achievement, and (c) that clusters are not independent from gender or home educational resources. Country differences in cluster numbers or composition were small.

The role of parents' literacy attitudes on children's reading achievement (PIRLS 2016)
Virginie Dupont, Dominique Lafontaine, Patricia Schillings, Stéphanie Géron, and Charlotte Dejaegher
(University of Liège, Belgium)

If positive parent reading attitudes are related to better pupil reading achievement, one wonders by which process this relation occurs. With PIRLS 2016 data for the French-speaking part of Belgium, our objective is to better understand the link between parents' reading attitudes and pupils' reading achievement. Our hypothesis is that parents' reading practices impact pupils' motivation which in turn is linked to pupils' reading achievement. Regression analysis shows that, taking into account socioeconomic status, reading self-concept and intrinsic motivation are mediators of this relation while instructional engagement is not.

The relationship between student engagement and achievement across countries within regions using latent class analysis
Renee Sze Leung Kwong and Greg Thomas Macaskill
(Australian Council for Educational Research, Australia)

One of the main purposes of student assessment reporting is to provide sufficient useful information to assist schools and policy makers to make decisions and implement policies to improve student learning. Student achievements in international studies are often reported as outcome measures by subgroups such as ESCS (Economic, Social, and Cultural Status), gender, and geographic locations. Frequently student performance is reported on up to two-level comparisons such as student mean performance by geographic location and school management. However, it does not necessarily follow that all students in any subgroup combination have the same characteristics and are facing the same challenges. This paper explores if student questionnaire indices, together with school background information, helps to characterize differences between countries and regions. By grouping students into homogeneous classes, it would provide not only a comprehensive idea of how the underlying characteristics of each group relates to their learning outcome but also provide an opportunity to compare results of individuals and schools in the same class. Data from the International Civic and Citizenship Education Study (ICCS) 2016 is used in this paper. Based on the similarity of culture, political climate, and geographic location, country data are grouped into three regions, Asia, Europe, and Latin America, with a focus on the Asia region. To look for meaningful grouping of student questionnaire indices by region, exploratory factor analysis (EFA) was first conducted on student questionnaire indices. The factors obtained showed some variation between the three regions. For example in the Asia region, four factors of nine variables are identified. The factors, represented by averages of their constituent variables along with school background variables, are modeled with Mplus using latent profile analysis (LPA) to obtain classes of students. The characteristics of these groups are then examined.

Session 2B (Symposium): Methodology
The future of international assessment? The promise and challenge of a multistage design in IEA studies
Chair: Leslie Rutkowski (Indiana University, US)
Discussants: Sabine Meinck[1] and Eugenio Gonzalez[2] (IEA[1]; ETS and IERI, IEA-ETS Research Institute[2])
Room: A403

Recognizing the advantages of computer-based assessment (CBA), IEA offered a digital component to PIRLS in 2016 and will transition to a digital platform for TIMSS in 2019. In particular, IEA cited efficiency gains, enhanced measurement, and innovative item types as advantages of a CBA. Given that dozens of highly heterogeneous educational systems take part in TIMSS, PIRLS, and other international large-scale assessments (ILSAs), a computerized platform offers a further advantage: the potential to include an adaptive element. To that end, multistage tests (MSTs) offer a number of opportunities and challenges in an ILSA setting. In response, the proposed symposium investigates several issues in this context. In the first of four papers, we consider the impact of an MST design on item parameter stability, whether probabilistic routing over and above merit routing has an effect on item parameter estimates, and how such a design translates into cross-country achievement distributions. In the second paper, we recognize the potential technological complexity associated with administering an MST by studying whether number-correct or item response theory methods are differentially effective at routing students to the correct stage and whether routing choices have an impact on achievement precision. A secondary aspect of this paper is how different module lengths impact routing decision accuracy, precision, and estimated achievement distributions. In the third paper, we investigate the interactive effects of an MST design and differential item functioning (DIF). In particular, we investigate whether DIF leads to misrouting and the performance of one measure to detect DIF. Finally, the low stakes nature of an ILSA makes motivation a possible source of construct irrelevant variance. To that end, the final paper in this symposium investigates how aberrant test taking behavior affects various aspects of an MST in an ILSA setting. Taken together, this series of connected papers offers insights into the viability and challenges of an MST design in future rounds of IEA studies and other ILSAs.

Parameter estimation stability and probabilistic routing in MST

Leslie Rutkowski and David Rutkowski
(Indiana University, US)

MST is a design that allows for limited adaptation of a test's difficulty to the proficiency of the examinee, in that adaptation happens at the module level, not the item level, as in fully computer adaptive testing (CAT). Here, module is defined as a group of items that always appear together in a block or cluster. A three-stage MST structure is referred to as a *panel*. Examinees begin in stage 1 with a core or routing module. Depending on the score in S11, examinees are routed to an easier or more difficult stage 2 module. Again, based on performance in the previous stage(s), the examinee is routed into a stage 3 module. Although the above description is limited to a single panel, an MST can be comprised of many panels. Routing students to the next module, based on their performance poses a potential problem in an ILSA context, where dozens of heterogeneous populations are tested. Less proficient countries will see easier items more frequently, while more proficient countries will see more difficult items more frequently, raising questions about item parameter bias. We consider whether a probabilistic mechanism that routes students to the next module, regardless of performance, can attend to bias without compromising the efficiencies of an MST. To that end, we simulate a design where 40 countries have generating proficiency values that range from highly proficient to not proficient. Then, we administer an MST design where routing decisions are based on merit only (merit routing), comparing estimated to generated item parameters. We then compare merit-based routing to probabilistic routing, where students have a 10 or 20 percent chance of being routed incorrectly regardless of performance. We evaluate item parameter estimates in terms of bias, root mean-squared error, and correlation with generating values. We repeat each condition 100 times for stability. Finally, we consider each design's impact on estimated proficiency distribution and country rankings.

Module length and routing methods in MST

Dubravka Svetina, Leslie Rutkowski, and David Rutkowski
(Indiana University, US)

In this paper, we study module length and routing methods (number correct vs IRT-based), emphasizing person parameter recovery under latent regression estimation. In particular, we examine how shorter vs. longer modules across stages can be optimized to obtain reasonable person parameter estimates. Secondarily, we examine how differing routing methods impact estimation accuracy across different panel constructions. We simulate a design and assemble tests based on test information functions, creating modules to target locations along the proficiency continuum that are representative of a typical ILSA. We manipulate four module sizes across all stages, either 6, 10, 12, or 20 items per module were selected in any condition, while maintaining a fixed number of 36 items for all simulees. This results in four designs, with balanced and imbalanced module lengths. Additionally, we examine two number-correct routing approaches (based on the last available module or based on cumulative performance from all previous modules) and two IRT-based approaches (Fisher's maximum information or expected a posteriori theta estimate). In conclusion, this study addresses two related aspects of the MST design that show promise yet are relevant to the ILSAs context—the panel construction/design features (panel characteristics, etc.) and routing methods—with the focus on understanding under which conditions we can recover person parameters with sufficient precision along the entire continuum.

Differential item functioning in MST

Montesrrat Valdivia, Leslie Rutkowski, and David Rutkowski
(Indiana University, US)

In spite of the potential benefits of MST, potential threats to test fairness remain. In addition to cultural, language, and geographic difference, other courses can include computer familiarity, internet connection quality, screen definition, the keyboard type, or test-takers with disabilities, potentially leading to bias. Where DIF exists in an MST, one consequence is that students might be misrouted to an incorrect module, effecting not only their performance on the DIF item, but also bearing consequences for future performance, as inappropriately difficult or easy items are administered in the next module. We evaluate through a simulation study whether misrouting occurs due to DIF and whether DIF at early stages of the assessment propagates into later errors of inference in terms of population and sub-population estimates. We also investigate differential effects when DIF occurs early vs. later in an MST. Then, as a means of detection, we evaluate CATSIB, which controls Type I error rates, has high statistical power, and gives less biased results. Although this method has been shown to work well in CAT, we contribute to the literature about CATSIB performance in MST. Further, there is no research that investigates DIF detection methods when DIF leads to misrouting. The design of the simulation will feature multiple populations, the characteristics of which are modeled after empirical ILSA data. We then build in DIF at each stage with DIF items that vary in magnitude and direction. CATSIB results will be evaluated in terms of Type I error and power of DIF detection. Misrouting results will be evaluated in terms of the proportion of students that are misrouted and how misrouting impacts achievement distributions.

On test taking motivation in MST

Yuan-Ling Liaw[1], Leslie Rutkowski[2], and David Rutkowski[2]
(University of Oslo, Norway[1]; Indiana University, US[2])

In TIMSS, student stakes are low, as performance has no personal consequences. As such, there might be little incentive for students to spend effort and take the test seriously, so students' test scores may not reflect their true proficiency, risking bias in country rankings. Such response behavior is known as *aberrant responding*, defined as an observed response pattern which deviates from the expected one based on a particular IRT model. Especially if aberrant responding is related to

proficiency levels, socioeconomic status, or other characteristics, it might produce biased achievement estimates and incorrect interpretation of the results. Therefore, it is important to understand the nature and magnitude of the threat posed aberrant responding, due to low motivation. However, there is limited understanding of (1) test-takers' aberrant behavior in a low-stakes exam; (2) the potential causes that drive students to respond aberrantly; and (3) the consequent effects on proficiency estimation and achievement ranking. Additionally, as some large-scale assessments have moved from linear test to adaptive testing (for example, PIAAC and PISA), no research has studied the threat posed by lack of test-taking motivation and effort and students' aberrant response behavior in MST. This paper will contribute to all four of these dimensions by using TIMSS 2015 data. First, we apply IRT-based person-misfit measures to detect aberrant behavior at the person level by analyzing response profiles of unexpected responses and summarize the proportion of "Omitted or invalid" and "Not reached" responses. We also correlate background characteristics and proficiency level with aberrant responses and we examine the effect of item characteristics on the probability of an aberrant response. Third, students who show highly aberrant behavior are excluded and effects on country rankings are estimated. Finally, a simulation study is conducted using simulated aberrant response behavior in an MST context. Unlike linear testing, examinees' scores or ability estimation on the first stage will determine how they are assigned to one of several modules in the subsequent stage in MST. If students respond aberrantly, the assigned module may not match a student's proficiency, leading to biased results. In the simulation study, different levels of aberrance for each defined aberrant response behaviors are manipulated. The accuracy of proficiency estimation and the robustness of country rankings will be further investigated.

Session 2C: ICCS
Civic attitudes in a changing world
Chair: Wolfram Schulz (Australian Council for Educational Research, UK)
Room: A405

Attitudes towards minorities: Changes from 2009 to 2016 using data from IEA's ICCS
Clara Beyer and Falk Brese
(IEA)

The United Nations Third Committee has been told by experts that migration is a "defining human rights issue of our time." Integral to the question of how migrants will incorporate into and interact with their new societies is an understanding of the manner in which host education systems receive migrant students. As reported by Côté and Erickson, minorities fare better in societies with higher levels of positive orientations called "tolerance." Tolerance can be seen as both a goal and a by-product of education. The study of education's effects on tolerance—and the unpacking of specific behaviors that occur in schools which correlate with differences in tolerant beliefs—is a timely one. This paper will analyze trends in the relationship between student civic engagement in school and attitudes towards minorities in countries participating in the International Civic and Citizenship Education Study (ICCS) 2009 and 2016. Drawing on modern incarnations of activity theory as a theoretical base, the authors will further unpack and analyze factors that are hypothesized to have a relationship with attitudes of students towards minorities—for example, student participation in school, engagement with social media, and discussion of political and social issues outside school. With a goal of clarifying specific activities performed within or outside of schools that have the strongest associations with increased levels of tolerance (within the ICCS framework, this can be quantified using reported attitudes towards equal rights), the authors will attempt to shed light on common factors across countries. Inherent to the discussion is the role of schools in the fostering of students' tolerant attitudes and behaviors.

Proximity to refugee centers and attitudes toward immigration
Diego Cortes and Mojca Rozman
(IEA)

Before fall 2015, refugees trying to reach western and northern European countries were predominantly using the Balkan Migration Route. This route starts in Greece and goes through Macedonia, Serbia, and finally connects to western Europe with the crossing of the Hungarian-Austrian border. On the 17th of October, Hungary's border closure shifted this migration route towards Slovenia. During the following months, Slovenia became the entry point into western Europe for refugees using this route. In this paper, we use the sudden temporary arrival of refugees to Slovenia as a source of plausibly exogenous variation to investigate the causal impact of refugee settlements on natives' attitudes towards immigration. In particular, we study the role that geographical proximity to refugee centers had on shaping these attitudes. Finally, we study whether geographical proximity favored any face-to-face interaction between refugees and natives, or whether other mechanisms were at work. First, we show that adolescents attending a school in a hosting local administrative unit (LAU) score about one-third standard deviations lower in the attitudes toward immigration scale, relative to those in a non-hosting LAU. Moreover, we show that students score higher on the scale as the distance between the school and its nearest refugee center increases. In order to interpret our results as causal, the identifying assumption is that in the absence of the refugee crisis, students in schools located close to a center would have followed the same trend in attitudes toward immigration as students in schools relatively far. We explore the validity of this assumption and show that the effect of refugee allocation is unrelated to several potential geographical differential trends. Finally, we study the mechanisms driving our results. We find no consistent evidence for a face-to-face interaction between natives and refugees.

Civic knowledge and civic self-efficacy: How do they contribute to the promotion of democratic attitudes?
Zsuzsa Blasko and Patricia Dinis Mota da Costa
(European Commission, Joint Research Centre, Italy)

As it is well established in the literature, civic and citizenship education provided in schools is moderately but positively associated with several non-cognitive civic outcomes including attitudes and behavior of students. Among the various features of the civic education process, the most commonly identified elements include formal learning process, democratic school experience, and (less frequently) involvement in community work. Research has repeatedly shown that each of these educational approaches have some influence on one or another democratic attitude—for example, support of citizenship values, openness towards diversity, or expected political behavior. Our paper contributes to the stream of research by analyzing the importance of civic and citizenship knowledge and civic self-efficacy in this attitude-shaping process. Theory considers both civic knowledge and civic self-efficacy as potential channels that might on the one hand be influenced by education and on the other hand might also affect students' attitudes. However, this mediating role has not yet been systematically tested. In our paper, we use ICCS 2016 data from the participating EU countries to investigate the process through which education affects students' perception of personal responsibility based citizenship, intended electoral participation, and openness towards equal rights for immigrants. Estimating a series of Structural Equations Models, we test the mediating role of both civic and citizenship knowledge and civic self-efficacy. Our findings show that schools indeed have some capacity to foster democratic attitudes—this is truer for some schools, and countries, than others. Civic knowledge and efficacy are complimentary assets, which mutually reinforce each other, still they only account for a small portion of the association between education and civic attitudes. This means that civic education has important direct effects that work irrespective of civic knowledge and self-efficacy. Indirect effects identified indicate that formal education not only increases civic knowledge but also contributes to the self-efficacy of the students, and in this way also promotes civic attitudes. Open classroom climate is a similarly important factor and a rather universal one, which again works directly as well as indirectly through both knowledge and self-efficacy. Finally, vocational work also shows some potentials, but it is interrelated with self-efficacy only and not with civic knowledge. We conclude our paper by discussing the limitations of the study that mainly arise from the difficulties of establishing causality from the cross-sectional data.

Explaining variation in teacher-reported student activities in the community
Falk Brese[1] and Julian Fraillon[2]
(IEA[1]; Australian Council for Educational Research, Australia[2])

The extent to which civic and citizenship education (CCE) is integrated in the curriculum varies considerably across education systems. This paper will explore variation within and across countries in one aspect of student-community engagement reported by teachers: their participation with students in civic-related activities in cooperation with external groups/organizations. We will investigate variation in teacher-led student co-operation with external groups/organizations within and between schools, and the contribution of teacher and school characteristics to variation. The paper will use data collected from over 94,000 Grade 8 (or equivalent) students and 37,000 teachers of Grade 8 (or equivalent) in the IEA International Civic and Citizenship Education Study (ICCS) 2016. We will complete a two-level multilevel modeling analysis, with teachers clustered within schools, using teacher and school-level predictors of teacher's reported teacher-led student co-operation with external groups/organizations. With these results, the paper can contribute to discussions of the degree to which the "whole school" approach is evident in one participatory aspect of CCE and how teacher- and school-level characteristics may contribute to teachers' likelihood to engage in these activities with their students. This proposed investigation is unusual in that we are using teacher-level data as an outcome variable to investigate the contribution of teacher and school contexts to the delivery of CCE within schools. Both the potential findings of the analyses and the methodological approach may support insights into CCE delivery and further research into CCE delivery within schools.

Session 2D: Engaging with IEA data
Roundtable on national experiences and use of IEA studies for teacher professional development
Chair: Josef Basl (Czech School Inspectorate, Czech Republic)
Discussants: Sue Thomson[1], Khaled Temsah[2], Elena Papanastasiu[3], Dana Pražáková[4], and Franck Salles[4]
(Australian Council for Educational Research, Australia[1]; Ministry of Education, UAE[2]; University of Nicosia, Cyprus[3]; DEPP, Ministry of Education, France[4])
Room: A401

The aim of the roundtable is to provide a space for presenting, sharing, and discussing specific examples of how individual IEA member countries use results, released test items, and so on, to support teachers, provide them with some feedback, and with some inspiration. The roundtable will focus on sharing experience related to the following questions. What benefits does participation of a country in IEA studies bring to its teachers? How do countries link national findings/materials with findings/materials from IEA studies to support their teachers? In which aspects could IEA studies do more to be even more useful for supporting teachers in specific educational systems? Experience from Australia, the UAE, Cyprus, and the Czech Republic will be shared within the roundtable. Panelists will focus for example on describing ways in which data as well as released test items from the Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS) could be utilized within an educational system for professional development purposes. A wide range of discussed policy measures will cover developing stakeholders' awareness about IEA studies, implementing international assessments into national policy indicators, or linking results of international assessments to school national evaluation to ensure implementation and accountability.

20 years of international large-scale assessments: How time series and trends can inform policymaking
Chair: Rolf Strietholt (Technische Universität Dortmund, Germany)
Discussant: Jan-Eric Gustafsson (University of Gothenburg, Sweden)
Room: A401

TIMSS 1995 may be regarded as the first international large-scale assessment (ILSA) of its kind. Norway took part in this study and a large number of other ILSAs over the next 20 years. In the national policy discussions, in key political documents, and through research, the comparative aspect of these studies has been the core element. After an initial emphasis mainly on comparisons with other benchmarking educational systems, over the last years we have witnessed increased attention to comparisons of the current state of affairs with the past, or, expressed differently, studies of change over time. Comparisons over time show that Norwegian students' performance declined in the first half of the period, followed by an approximately equally large growth in the last half. In addition, over time Norwegian students report more positively on questions relating to students attitudinal attributes, the learning environment, and school climate. Over time, the Norwegian school system (and other Nordic systems) has also been reported to be characterized by relatively small differences between schools, between genders (except for reading), and between students with different home backgrounds. However, recent research has indicated that this particular aspect of the Nordic countries may be in flux. This symposium presents some features of the development of these core characteristics. Three papers are included with studies of changes in students' achievement, their motivation, and in equity, respectively. Although the papers exclusively present analyses from the Norwegian context, the intention is to stimulate a more general and broad discussion about how time series and trends can inform policymaking. The overarching question framing the discussion is the following: What, if anything, could be done with the designs of ILSAs to optimize the possibility to study change over time? Alternatively, to phrase this more as a provocative challenge, is it possible to maintain stable time series, and at the same time ensure that the studies are perceived to be relevant for the current and future educational reality?

Predicting change in mathematics achievement in Norway over time
Rolf Vegar Olsen and Sigrid Blömeke
(University of Oslo, Norway)

The increase in Norwegian Grade 8 students' mathematics score TIMSS 2003 to 2015 was substantial (25.5 points). In fact this change is equal to the difference between Norwegian eighth and ninth graders in 2015 (Norway participated with both grades). In this talk, we will present and discuss results from an analysis where this change in achievement is related to changes in 22 context variables representing two broad characteristics: 1) characteristics of the student cohorts; and 2) characteristics of the schools and instruction. By applying Oaxaca-Blinder decomposition (OBD), changes in averages for the context variables and their relation to achievement in the period is used to evaluate the change in mean mathematics achievement. The analyses revealed that almost all included variables changed significantly between 2003 and 2015, and most of them in a direction one would expect to influence achievement in a positive way: Students value mathematics more highly, have higher self-confidence with respect to mathematics, and they enjoy schooling more in 2015 than in 2003. Teachers have higher teacher education degrees, are more often specialized in mathematics or mathematics education, and feel that it is safer at school. Furthermore, teachers, students, and parents report that they are more ambitious academically. During the same time, a few variables changed significantly towards a direction one would expect to influence achievement in a negative way: The proportion of students with a migration background almost doubled, and the proportion of students with lower cultural capital (as assessed by the number of books at home) increased. Correspondingly, the contribution from these different developments in the OBD cancelled out in the total model. However, the components with negative contributions are largely outside the control of the educational system, and although the model is strongly affected by unobserved variables, our interpretation is that the improved mathematics performance in Norway is partly accounted for by changes in the school context and students affective-motivational attributes.

Has equity changed in Norway over the last decades?
Trude Nilsen, Julius Björnsson, and Rolf Vegar Olsen
(University of Oslo, Norway)

Equity is an important aspect of the school system, and differences in achievement related to home background are small in Nordic countries seen in an international perspective. However, previous research indicates that this might be changing. This study examines changes in equity in Norway during the last 20 years using data from both TIMSS (Grade 4 and 8) and PISA (15-year-old students) for the whole period from 1995 to 2015. Multi-level regression models were estimated using Mplus. Different measures of equity were included. Changes over time of total variances in mathematics achievement, and variance decomposition at the student and school level were investigated, as well as variance explained by SES. In addition, the changes in the strength of the relation between SES and achievement over time were estimated. Findings show that the differences in achievement (variance) at the school level increases throughout the period, in two of the three analyzed age/grade cohorts, but decreases in the last cycle of the studies. Concurrently the variance at the student level generally decreases, but increases between the last two cycles. When comparing results from PISA and TIMSS, the relationship between student home background and achievement does not appear to have a consistent tendency throughout the period. With the exception of Grade 8 at the school level, where the strength of this relation has doubled between 1995 and 2015, the relationship between student home background and achievement increases slightly through the period for TIMSS while it is unchanged in PISA. The results are discussed in light of how different measures of equity may generate different and often diverging results. The implications for educational policy are discussed.

Twenty years of science motivation mirrored through TIMSS: Example of Norway
Hege Kaarstein and Trude Nilsen
(University of Oslo, Norway)

Student motivation is important for recruitment to further STEM education and carrier. Over the last decades, Norway has allocated resources for recruitment to STEM, especially for girls, making it important to explore how students' motivation has changed across time. Using data from TIMSS, this paper explores the changes in Norwegian students' motivation and the relations to achievement in science across time (N=43,366). Multi-group structural equation models as well as measurement invariance analyses of comparability across gender, grades, and cycles were conducted in Mplus. Findings indicate an increase in motivation and in the strength of the relation between motivation (self-concept, intrinsic, and extrinsic) and achievement across time for both grades, and higher motivation in favor of boys. While all motivational constructs were comparable across time, self-concept was not comparable across gender, and no construct was comparable across grades. These findings have implications for policymaking and teaching practices in science classrooms in Norway.

Session 3B: Methodology
Methodological issues in large-scale assessments: Challenges and solutions (1)
Chair: Julius Björnsson (University of Oslo, Norway)
Room: A403

Achievement gaps between immigrant and native students over time: Evidence from 20 years of TIMSS
Ebru Erberber, Yifan Bai, Yuan Zhang, Frank Fonseca, Yemurai Tsokodayi, and Marissa Hall
(American Institutes for Research, US)

With large numbers of immigrants of diverse backgrounds entering various destination countries, immigration status has not only become an important issue on the political agenda within societies, but also a critical aspect in addressing educational inequality. It becomes important to understand, within and across countries, the differences in educational achievement between native students and students from immigrant backgrounds (hereafter referred to as "immigrant students"). This study examined the proportion of immigrant students and their achievement compared to native students in 16 education systems using up to two decades of data from the Trends in International Mathematics and Science Study (TIMSS). Results suggested that most education systems observed an apparent increase in the proportion of immigrant students, while other education systems experienced limited increase or decrease to varying degrees over time. Moreover, regarding achievement gaps between immigrant and native students, three general trends emerged, with education systems showing either apparent increasing, decreasing, or reversing gaps. This general trend in most education systems towards increases in immigrant student populations underscores the need to understand these populations and the factors that affect their academic achievement. Further research should explore how differences in the characteristics of immigrants, and the policies of the countries they reside in, influence the outcomes of these immigrant students.

Aiming for measures of long-term trends in Grade 8 mathematics: Bridges from SIMS to TIMSS 2015
Monica Rosén, Erika Majoros, Stefan Johansson, and Jan-Eric Gustafsson
(University of Gothenburg, Sweden)

International comparative assessments on student achievement provide a large body of data spanning about five decades. International large-scale studies are constructed to assess country-level change. A coherent understanding of the international trends in educational outcomes is strongly needed and suggested by numerous previous studies. Given that education is rather stable, investigating these trends requires long-term analysis. There is previous research on linking large-scale studies over time for cognitive outcomes related to reading ability. However, there are no previous attempts to link the background questionnaire items measuring non-cognitive constructs over time. The purpose of the present study is to investigate the possibilities for the long-term scaling of both cognitive and non-cognitive outcomes in international large-scale assessments in mathematics. The present study, which is part of a larger project with the aim of studying long-term trends, uses data from Hungary and Sweden, as these countries are among those participating most frequently in the assessment cycles on mathematics. The analysis includes the construction of an inventory of items measuring Grade 8 achievement and attitude in studies on mathematics conducted by IEA. The second part of the study attempts to create common scales across time, by linking six studies from 1980 until the present. The document analysis indicates that there are possibilities for linking these studies in both types of outcomes, as there are a number of overlapping items. As there are no previous attempts to link items measuring attitude over time, this section of the study is by necessity highly explorative.

Sensitivity of the RMSD for detecting item-level misfit in low-performing countries
Jesper Tijmstra[1], Yuan-Ling Liaw[2], Maria Bolsinova[3], Leslie Rutkowski[4], and David Rutkowski[4]
(Tilburg University, the Netherlands[1]; University of Oslo, Norway[2]; ACTNext, US[3]; Indiana University, US[4])

While the root mean squared deviation (RMSD) is a popular statistical measure for evaluating country-specific item-level misfit (i.e., differential item functioning) in large-scale educational assessment, this paper shows that its sensitivity to detect misfit may depend strongly on the proficiency distribution of the considered countries. Specifically, items for which most respondents in a country have very low (or high) probability of providing a correct answer will rarely be flagged by the RMSD as showing misfit, even if very strong differential item functioning is present. With many large-scale educational assessment initiatives moving towards covering a more heterogeneous group of countries, this raises issues for the ability of the RMSD

to detect item-level misfit, especially in low-performing countries that are not well-aligned with the overall difficulty level of the test. This may put one at risk of incorrectly assuming measurement invariance to hold, and may also inflate estimated between-country difference in proficiency. The degree to which the RMSD is able to detect differential item functioning in low-performing countries is studied through simulation.

A new powerful validation procedure to guarantee comparability in international attitude and test scales: 'Alignment'
Ingrid Munck
(Gothenburg University, Sweden)

This paper serves to introduce a new flexible and powerful validation procedure; 'Alignment', a method which automates the control and testing of survey quality and comparability in ILSAs across multi-national, multi-regional, and multi-cultural contexts (3MC). This method offers the principal investigator a protocol of the measurement invariance status reported for each different context/group estimated on the basis of respondents' answers to the items in a scale. The Alignment method is especially useful for international survey organizations who wish to raise, and assure, the quality of reported group mean differences and trends in scale scores. By introducing additional potential bias components, like gender or age in the design/grouping, the non-invariance related to different sources of measurement errors can be explored and tested. The Alignment procedure implemented in Mplus reports on the amount of invariance and non-invariance for items, as well as for every model parameter and group in a common metric. Findings about significant non-invariance will identify total survey error (TSE) components that need to be addressed in order to achieve 3MC survey comparability. This method provides a refining of scales for pooled data from a large number of groups, using the same instrument. The Alignment approach is also able to generate aligned factor scores for all individuals in the data set, despite the presence of significant non-invariance in some groups. This is a flexible way to keep all observations in the analysis and to report results for all participating respondents. A full-scale application of the Alignment method is reported about support for immigrants' rights in a pooled data set from the 1999 IEA Civic Education Study (CivEd) and the 2009 IEA International Civics and Citizenship Education Study (ICCS). The alignment analysis of adolescent measurement invariance across 92 groups (country by cohort, by gender) found that a five-item scale was statistically well grounded for unbiased group comparisons despite the presence of significant non-invariance in some groups. The misfit could be located in just a few groups, and there were only marginal effects when these were incorporated into the reported results. Further cross-validation and replication of other scales and tests (including using new waves of IEA data collection, for example, the 2016 ICCS wave) will demonstrate how the potential of the Alignment technique can be adopted as a procedure for validation of scales in ILSA to ensure best possible measurement comparability.

Session 3C: ICCS
Social media and civics culture: Evidence from ICCS 2016
Chair: Julian Fraillon (Australian Council for Educational Research, Australia)
Discussant: Maria Magdalena Isac (University of Groningen, the Netherlands)
Room: A405

Social media harm institutional trust and thereby civic culture in Europe
Mikael Sandberg
(Halmstad University, Sweden)

Trust in institutions and political participation are both essential for democracy. This paper proposes modeling trust in institutions among 14-year-old students in schools comparatively using ICCS 2016 data from 13 European countries. Data include more than 54,000 students in their eighth grade of schooling. Based on previous studies, we hypothesize to find that social media harm trust in institutions and therefore negatively affect civic and democratic participation. Trust and civic culture systems are studied using structural equation modeling (SEM). Results indicate that trust in social media indeed negatively influences trust in institutions in almost all European countries, which in turn reduces expected political and electoral participation of students.

Young people's use of social media and internet for civic engagement in 24 countries
Wolfram Schulz and Tim Friedman
(Australian Council for Educational Research, Australia)

The use of digital technologies for citizenship engagement has become more widespread over recent years and has led to the conceptualization of 'digital citizenship' as an emerging feature of citizenship participation in societies at the outset of the 21st century. This paper uses data from the recent ICCS 2016 survey to investigate the scope of using social media and the internet for civic engagement among lower-secondary students in 24 countries. It analyzes its relationship with context factors and other forms of engagement, and models the influence of a range of contextual and other related variables on variation in students' use of social media for civic engagement.

School, church, and media: Explaining differences in civic knowledge in the bilingual school system of Estonia
Triin Lauri[1], Kaire Põder[2], and Anu Toots[1]
(Tallinn University, Estonia[1]; Estonian Business School, Estonia[2])

This study investigates civic and citizenship education in the unique post-Communist context of Estonia's bilingual education system. Estonia has a bilingual school system where there are Estonian and Russian language schools running in parallel.

While Estonian language school students are ranked very high in international comparisons, there is a significant difference between the achievement of Estonian and Russian language school students across different measurements. We claim that this ethnic achievement gap in the performance of civic and citizenship knowledge (CCK) is mostly explained by behavioral and attitudinal factors that are moderated by the school language. We rely on hierarchical modeling to capture the embedded data and aim to explain how the different layers (school and student level) interact and impact CCK. We show that behavioral and attitudinal explanatory factors (such as a less open classroom climate, differences in trust in media, and impact of religion on civic life) explain the gap in CKK between Estonian and Russian language schools.

Session 3D: Engaging with IEA data
Symposium on *IEA Compass: Briefs in Education* series
Chair: David Rutkowski (Indiana University, US)
Discussant: Dirk Hastedt (IEA)
Room: A220

This symposium will focus on findings from the *IEA Compass: Briefs in Education* series. Based on IEA studies, the *IEA Compass* series addresses issues of interest to a broad range of educational stakeholders, especially, but not limited to, those involved in influencing educational decision and policymaking. Each publication in the series aims to connect study results to recurrent and emerging questions in education at the international and national levels. The symposium is divided into three sections. The first section will consist of a short introduction by the IEA Executive Director Dr Dirk Hastedt. Dr Hastedt's opening remarks will focus on the goals of the IEA's series and how he feels educational stakeholders around the world can use the briefs to inform and improve their own educational systems. Following Dr Hastedt's remarks, four authors will respond to specific questions about their work. There will be no PowerPoints but authors will share one table or graph from the brief on the screen during the talk. After the presentations by each author, Dr Rutkowski will provide general remarks concerning the briefs and presentations. He will also discuss the future of the series and discuss upcoming topics to be covered. The remainder of the time will be open to audience questions and comments.

PIRLS for teachers: Making PIRLS results more useful for practitioners
Therese Hopfenbeck and Jenny Lenkeit
(University of Oxford, UK)

The IEA's Progress in International Reading Literacy Study (PIRLS) is a valuable resource for educational researchers and policymakers, but not well understood by teachers. To improve the impact of PIRLS data on pedagogy in the classroom, researchers and local teachers worked together on Oxford University's PIRLS for Teachers project. This project aimed to provide teachers in England with good guidance on interpreting and using knowledge from PIRLS to improve their own teaching of reading in primary schools. Outcomes included podcasts, posters, and videos supporting best teaching practices. Education systems participating in PIRLS are advised to involve teachers in the interpretation of the PIRLS data to ensure educators can make an informed contribution to the national dialogue. As well as developing podcasts, posters, or similar materials for practitioners, workshops or practitioner conferences where teachers can interact with researchers foster valuable dialogue on the best use of the data collected by PIRLS.

Preparing the ground: The importance of early learning activities at home for fourth grade student achievement
Sabine Meinck and Agnes Stancel-Piątak
(IEA)

Enhancing parental and child engagement with stimulating pre-school activities at home can contribute to improved child development and school outcomes, in particular when children come from socioeconomically disadvantaged backgrounds. Drawing from TIMSS and PIRLS data our analyses suggest that early learning activities have some effect on student achievement in later schooling. Further, in some education systems, parental engagement with children on stimulating learning activities is unequal across families from different educational backgrounds. Moreover, better educated parents tend to support their child's development with greater frequency and intensity than parents with financial and educational limitations. In this regard, policy action could be directed toward enhancing the nurturing environments of young children. Further, with our findings and using results from the regional Project on Child Development Indicators (PRIDI), we also argue for an approach to raise awareness of the lifelong impact of pre-school education, and to offer parents or caregivers opportunities to learn how to be more responsive to their children's needs. Finally, our analysis indicates that interventions would be most beneficial for children from families where the parents have lower levels of education.

Is democracy overrated? Latin American students' support for dictatorships
Andres Sandoval-Hernandez
(University of Bath)

Results from ICCS 2016 show that 69 percent and 65 percent of the students in the participating Latin American countries would support a dictatorship as a form of government if it brings order and security or if it brings economic benefits, respectively. In this brief, we look closer at these findings in order to identify possible explanations and potential policy implications. Our results show that, this apparently worrisome figure should be taken into perspective since these percentages have decreased in comparison to the results from the previous round of ICCS. Further, we find that higher levels of support for dictatorships are negatively associated with students' civic knowledge and with the frequency with which

students talk with their parents about political issues. These relationships, however, are mediated by students' trust in institutions. We discuss the potential role of families and formal education in instilling democratic values in future citizens, but make it clear that not all depends on schools and families. We suggest that failing democratic systems, where citizens do not trust their institutions, need to improve transparency, eradicate corruption practices and the delivery of services to the citizenry in order to promote the general improvement of support for democratic values.

No one likes a bully: How systematic is international bullying in fourth grade?
Leslie Rutkowski and David Rutkowski
(Indiana University, US)

TIMSS is a valuable resource to inform better understanding of bullying trends in general, as well as revealing the relationship between bullying and academic achievement, both within and between countries. The TIMSS study is unique in that over the past 20 years it has administered a bullying scale to fourth and eighth grade students from around the world. The responses that students provided to the TIMSS 2015 student questionnaire provide a unique perspective on bullying around the world. Our findings show that bullying is not isolated to one country. Rather, bullying is an international phenomenon that spans cultures and economies and begins at an early age. In the brief, we provide evidence of a strong international association between bullying and mathematics achievement at the fourth grade confirming that bullying begins at an early age. Based on our analysis presented in the brief we recommend that educational policymakers should address the issue of bullying by carefully examining their own contexts and using interventions that are proven to work best in a given setting. Nevertheless, analysis of a large-scale assessment such as TIMSS, with many countries and representative samples, demonstrates that bullying victimization is happening across a wide range of heterogeneous countries, regardless of geography, dominant race/ ethnicity, language, culture, and economic development. Further, irrespective of rank, in the majority of countries there were lower levels of achievement where bullying was prevalent. Finally, the longitudinal design of TIMSS, which assesses fourth and eighth grade students every four years, offers policymakers an opportunity to use TIMSS results to investigate bullying trends over time, within and between countries.

Session 4A: TIMSS and PIRLS
Explaining the gaps: Factors of student achievement
Chair: Falk Brese (IEA)
Room: A401

Differences in students' and teachers' characteristics between high and low performing classes in Slovakia
Samo Varsik
(Ministry of Education, Science, Research, and Sport of the Slovak Republic, Slovak Republic)

The design of the PIRLS 2016 survey can be utilized to explore many characteristics about schools, teachers, and students in Slovakia which were previously not known. PIRLS 2016 can also be used to examine which characteristics differ between high and low performing classes in Slovakia. Generally, variables that were broadly associated with the socioeconomic background of students or schools yielded statistically significant differences. However, very few significant differences in teaching methods or teachers' characteristics were found. Using hierarchical linear models it was found that socioeconomic background is one of the most important factors in explaining students' performance. Teachers' characteristics or teaching methods were not significant after controlling for socioeconomic status. The impact of students' confidence on their performance is one avenue that should be explored further, in particular for students from poor socioeconomic backgrounds. The marginal effect of students' confidence is higher for students from lower socioeconomic backgrounds. This suggests that if confidence in reading could be strengthened for students from lower socioeconomic backgrounds, these students could see significant performance gains.

Reading comprehension growth from PIRLS Grade 4 to Grade 6
Bieke De Fraine, Koen Aesaert, and Jan Van Damme
(KU Leuven, Belgium)

Flanders' (the Dutch-speaking part of Belgium) average score declined significantly between PIRLS 2006 and PIRLS 2016 (minus 22 points). In order to clarify this decline, the students that participated in PIRLS 2016 were reassessed when they were in Grade 6 (2018). The longitudinal data of this study enable us to expand the regular trend analyses of PIRLS 2016 with results describing learning gains on reading comprehension at the student level, at the school level, and at the level of the educational system.

Nordic students' achievement and school effectiveness in TIMSS 2015
Marie Wiberg and Ewa Rolfsman
(Umeå University, Sweden)

The effectiveness of schools is often defined by students' success in core subjects and a basic assumption within this research field is that students' success are related to school factors, which can be categorized into context and climate variables. A key actor is the principal, since he or she is ultimately responsible for the school and thereby ultimately the students' success. The overall aim of this study was to identify factors in the school environment that are associated with students' mathematics achievement in the Nordic countries. We focused on a comparison of Sweden and Norway, as they both participated in TIMSS 2015. Effective schools were identified by using average mathematics achievement and

constructed regression models with the student factors and variables as covariates. Schools were regarded as high effective if they were in the top third in their country in mathematics achievement, mid effective if they were in the middle third, and low effective if they were in the bottom third in their country. School factors were constructed based on the school questionnaire answered by the principals. Multilevel analysis was used to separate the school effects from the students' home background. The overall results indicate that few school factors associated with student success were found, in particular in Sweden where school factors of importance are mainly related to the student body regardless of the different school types (i.e. low-, mid-, or high-effective school). School factors of importance for student success in Norway appear mainly to be related to the school climate. The finding for Sweden is in line with previous studies. However, this is a result that requires some thought in light of the recent development in Sweden regarding the rise of school segregation and the Swedish long-term goal of encounter threats to individuals, equal opportunities, and access to education, and future plans on educational investments.

Birth month and mathematics performance relationships in Norway
Annette Hessen Bjerke, Elisabeta Eriksen, Andre Rognes, and Bjørn Smestad
(OsloMet-Oslo Metropolitan University, Norway)

Due to a fixed school start in August of the calendar year of pupils' sixth birthday, the age span in one class in Norway is up to twelve months. This contributes to the mathematics performance span in the early years and later. In this paper, we investigate the relationship between birth month and mathematics performance (the relative age effect) by paying attention to the content and cognitive domains addressed in TIMSS 2015. In Norway, TIMSS 2015 included four cohorts, enabling a comparison between fourth and fifth grades, and eighth and ninth grades. We find significant correlations between birth month and mathematics performance in all content and cognitive domains for Grades 4, 5, and 8. For ninth grade, there are no significant correlations in any content or cognitive domains. There are no significant differences in age effect between fourth and fifth grade or between eighth and ninth grade.

Session 4B: Methodology
Methodological issues in large-scale assessments: Challenges and solutions (2)
Chair: Monica Rosén (University of Gothenburg, Sweden)
Room: A403

Measurement invariance of mathematics attitudes scale across the Gulf Cooperation Council countries
Abeer A. Alamri
(National Center for Assessment, Qiyas, Saudi Arabia)

Despite the wide use of attitudes scales, measurement invariance of mathematics attitudes scales has rarely been investigated. Using eighth-graders data from TIMSS 2015 Student Context Questionnaires of the Gulf Cooperation Council countries (GCC), the Mathematics Attitudes Scale (MAS) factor structure, and measurement invariance were examined. Based on the findings, MAS psychometric properties were sound, a strong invariance was established, and latent means are estimated and compared. Mean differences may provide greater justification for a potential link between attitudes and performance. Educational psychologists, policymakers, and practitioners in GCC countries could have better communication with students by understanding the potential effect on students' performances and exerting effort for improvement.

Item parameter equivalence in TIMSS 2015
Huseyin Husnu Yildirim[1], Selda Yildirim[1], and Norman Verhelst[2]
(Bolu Abant Izzet Baysal University, Turkey; Eurometrics[2])

This current study investigates item parameter equivalence in mathematics and science across forty-four participants in TIMSS 2015. The responses of 280,130 students to 209 mathematics and 215 science items are studied. Item discrimination and difficulty parameters are estimated separately in each of the participants. Results demonstrate violations of the assumption of item parameter equivalence across the participating countries in TIMSS 2015. For example, the Pearson correlation coefficient in mathematics item difficulty estimates fell down to 0.29 between Kazakhstan and Sweden. Results also show that item parameters are relatively more similar across the countries from similar regional, cultural, and developmental backgrounds. Developed western countries, Arabic countries, and East Asian countries form clear clusters in multidimensional scaling plots. Fluctuation in item parameters across the countries manifests itself at item-level differential item functioning analyses, as expected. However, when item categories are considered, the effect is barely evident in countries' deviation performances calculated via profile analyses. Measures that could be taken to address the item parameter equivalence problem, which is perhaps inevitable, are discussed.

Performance differences between TIMSS's paper-pencil test and tablet test of Dutch Grade 4 students
Eva R. Hamhuis, Martina R.M. Meelissen, and Cees A.W. Glas
(University of Twente, the Netherlands)

Over the past few decades, the educational use of ICT has become an important part of primary education in the Netherlands. In Dutch primary schools computers and tablets are not only frequently used for instruction and learning but are also used for the assessment of students. Research has indicated that the specific characteristics of a paper-pencil based test (PBT) and those of a computer based test (CBT) might have an impact on student performance. Research has also suggested that the occurrence of a mode effect depends on the kind of device used, the subject that is assessed, and student

characteristics such as gender or reading ability. The participation of the Netherlands in the Equivalence Study of eTIMSS 2019 in the spring of 2017, offered the possibility to explore the mode effects of a low-stake, summative assessment in mathematics and science. In total 532 Dutch, Grade 4 students were assessed with a paper-pencil test and a tablet test. Item response theory (IRT) was used to explore potential mode effects. The study revealed no mode effect for Dutch, Grade 4 students. For both mathematics and science, there were no differences found between the student ability scales and item parameters. However, Dutch girls slightly outperformed boys on the tablet test. No differences were found for the different levels of reading demand between the two test modes.

How to use the world as a laboratory: Teachers' role in enhancing equity

Trude Nilsen[1] and Jan-Eric Gustafsson[1,2]
(University of Oslo, Norway[1]; University of Gothenburg, Norway[2])

This paper examines the relation between teacher quality and equity, and demonstrates an approach to "using the world as a laboratory" by including all countries while taking into account cultural differences. The aim of most educational systems is to increase educational equity. There is hence a need to identify factors that may reduce the strong effect of student home background (or socioeconomic status, SES) on student outcome. The closest proximal factors to students, and key to their learning outcome, are teacher quality and their instruction. Yet, few have linked these to equity. Even fewer have investigated this across countries. Although ILSA data provide an excellent opportunity to study questions of equity, there are major challenges related to the large cultural differences between the participating countries. The present paper suggests a three-step analyses procedure that may reduce these challenges and sets out to examine whether teacher quality and their instruction may weaken the relation between SES and mathematics achievement, and thus reduce inequities among students. Including all countries who participated in Grade 8 in 2015 (N=220, 609), a three-step approach was undertaken. In step one, hierarchical cluster analysis was employed to identify countries that cluster together in terms of SES and student achievement. The Wards method with squared Euclidian distance was chosen. In step two, measurement invariance analyses were done to examine to what extent measures were comparable within the clusters. In step three, two-level random slopes structural equation models were estimated in Mplus, to investigate whether teacher qualifications and feedback practices may moderate the relation between SES and achievement. Preliminary findings identified four clusters of countries, and our measures were either scalar or metric invariant across countries. The comparability was hence sufficient to proceed to the moderation analyses in step three. Findings from step three indicated that participating in professional development as well as teachers' feedback practices weakened the relation between SES and achievement, thus promoting equity among students. The findings have implications for teacher education, educational researchers, and applied methodology. Furthermore, the results should be of interest to educational policy, as they demonstrate that teachers may make a difference in reducing inequity through their qualifications and instruction.

Session 4C: ICCS
Regional analysis of ICCS data
Chair: Andres Sandoval-Hernandez (University of Bath, UK)
Room: A220

Change or stability? Civic knowledge, classroom climate, and expected political participation in Latin American schools

Daniel Miranda and Juan Carlos Castillo
(Pontificia Universidad Católica de Chile, Chile)

The participation of the majority of citizens in public spaces plays a relevant role in both the strength and legitimacy of democracy. However, it is clear that achieving the participation of the majority of citizens is still a relevant challenge, especially in times characterized by political disaffection and political inequalities. It is expected that early socialization within the family, and particularly at school, should aim to provide the tools that allow future citizens to become successfully involved in political life. The present paper aims to analyze the possible impact that schools can have in mitigating the effect of parents' socioeconomic status on students expected electoral participation, focusing on two variables: civic knowledge and classroom climate. Additionally, it is aimed to explore the stability or change of the school socialization processes described. The analyses are based on a series of multilevel models using Latin American countries data of the International Civic and Citizenship Education Study (ICCS) 2009 and 2016. The results support the influence of students' socioeconomic background on expected electoral participation. Furthermore, civic knowledge and classroom climate show a positive and similar influence on students' expected participation. These findings support previous studies whereby an open climate within the classrooms allows students to acquire higher levels of civic knowledge, and furthermore it influences their expected political participation, becoming a powerful aspect to take into account when mitigating the political participation gap among students from different socioeconomic backgrounds. Civic knowledge and climate, although through different channels, mitigate background effects on participation in the countries considered. Nevertheless, there are several aspects that still need to be considered. For instance, the link between countries curricular guidelines and results in knowledge, as well as the differential influence of background variables on expected participation. Regarding future research, in a following study the idea is to extend the model to the complete ICCS database, in order to compare Latin America with the rest of the countries.

Young people's support for authoritarianism, trust in institutions, and expected political participation in five Latin American countries

Wolfram Schulz
(Australian Council for Educational Research, UK)

Research in recent years has indicated a concerning lack of commitment in support for democratic government across Latin American countries, both among adults and young people. This paper investigates the extent of lower-secondary students' support for authoritarian government practices and its relationships with trust in civic institutions and expected political engagement. The analysis also focuses on the role of civic learning factors with regard to students' perceptions of public institutions. The results confirm strong negative associations of civic knowledge with authoritarian orientations as well as with trust in civic institutions and more active forms of expected political participation.

Latin American students' attitudes toward the use of violence

Citlalli Sanchez-Alvarez
(Instituto Nacional para la Evaluación de la Educación, Mexico)

The aim of this study is to examine differences in attitudes toward the use of violence amongst students of the five Latin American countries that participated in the International Civics and Citizenship Study (ICCS) in its 2016 cycle: Chile, Colombia, Dominican Republic, Mexico, and Peru. The study examined the relationship between school factors and students' acceptance and endorsement of violence in a sample of 25,319 eighth grade Latin American students using data from the International Civic and Citizenship Study (ICCS) 2016. It explores the ways in which students' contextual variables, attitudes, perceptions of their schools' learning environment, and school context contribute to their attitudes toward violence. ICCS 2016 databases for Chile, Colombia, Dominican Republic, Mexico, and Peru were used to run multilevel regression analyses using different variables at student and school levels, including variables from the Latin American student questionnaire. Results from the regression models are described.

What you know and what you believe: Explaining tolerance of corruption in Latin American students

Diego Alonso Carrasco Ogaz[1], Robin Banerjee[2], Ernesto Treviño[1], and Cristóbal Villalobos[1]
(Pontificia Universidad Católica de Chile, Chile[1]; University of Sussex, UK[2])

The promotion of citizenship is an aim of many educational systems, but we lack a detailed understanding of how these efforts translate into civic attitudes. We proposed a model whereby links between socioeconomic status (SES) and tolerance of corruption are explained by civic knowledge, authoritarianism, and open classroom discussion. Using data from the Latin American samples of the International Civic and Citizenship Study 2009 (ICCS 2009), our specified model fits the data well, accounting for 42 to 53 percent of the variance in eighth grade students' tolerance of corruption across six countries. Students from disadvantaged backgrounds are more tolerant of corruption, yet this effect is fully mediated by their civic knowledge. Students with higher civic knowledge were less tolerant towards corruption practices, partially explained by lower endorsement of authoritarianism. Additionally, in four out of the six countries, this indirect effect was found to be larger when students were exposed to more open classroom discussion.

Session 4D: Engaging with IEA data
Innovative approaches to analyzing IEA data
Moderator: Sabine Meinck (IEA)
Room: A405

This symposium focuses on three innovative areas for analysis of IEA data: trend analysis, analysis of the combined PIRLS and ePIRLS datasets, and use of R to analyze data and estimate student achievement. The purpose of this symposium is to expose attendees to novel approaches for analyzing IEA data. The first two presentations focus on analysis of IEA trend data. Generally, researchers look at one cross-section of IEA data (for example, PIRLS 2016) and examine correlations between contextual factors and student achievement for one assessment cycle. Building on Gustafsson's trend analysis approach, the first presentation explores examining correlates of PIRLS reading achievement across the first four cycles of PIRLS. The second presentation provides trend results of socioeconomic achievement gaps in 13 education systems over a 20-year-period and details methods-related issues involved with analyzing trends across multiple cycles of IEA data. The next two presentations explore analysis possibilities using the combined PIRLS and ePIRLS datasets. The third presentation uses the combined datasets to explore student access and use of computers both at home and in the classroom, and the relationship between each of these factors and ePIRLS student achievement. The fourth presentation builds on this by examining to what extent computer use and access explain ePIRLS achievement after controlling for student offline reading ability (their PIRLS achievement). The final presentation summarizes the current capabilities of the EdSurvey package in R—a package designed to enable the analysis of large-scale assessment data in R (including IEA data)—and discusses two innovations of the package: multiple imputation and direct estimation. Through analysis using simulated data, the results of multiple imputation and direct estimation analysis are explored and compared to results generated by other methods.

Country-level trends in early literacy activities: Predicting gains in PIRLS reading?

Martin Hooper
(American Institutes for Research, US)

IEA data is typically analyzed cross-sectionally—we correlate contextual factors with student achievement in one cycle (for example, PIRLS 2016). Gustafsson (2007) suggested that we can gain new insights by examining correlates of country-level trend data—to what extent there is a correlation between trends in background data and trends in student achievement data. With PIRLS assessments completed in 2001, 2006, 2011, and 2016, this paper examines long-term trends in early

literacy activities and extends Gustafsson's approach across multiple cycles analyzing to what extent country-level increases in early literacy activities over time are associated with country-level increases in PIRLS reading achievement. The data are analyzed for the 16 countries and two benchmarking entities that participated in the four PIRLS cycles.

Analyzing trends in socioeconomic inequality and educational outcomes: Results and methodological strategies
Yifan Bai
(American Institutes for Research, US)

This presentation first reports on trend results for TIMSS socioeconomic achievement gaps analysis across 13 education systems over a 20-year-period, and then details methodological issues and strategies for measuring trend using IEA data. While the overall positive association between family socioeconomic status (SES) and student achievement is well documented in the literature, the magnitude of this relationship is contingent on social contexts and is expected to vary by education system. Research is limited on how such associations differ internationally and how the strength of these relationships has changed over time. This study fills this gap by reporting on changes in achievement gaps due to SES over time in these education systems. Second, the presentation discusses methods-related issues involved in analyzing trends across multiple cycles of IEA data, using examples and lessons learned from the 20-year trend analysis. The examples provided show steps taken to identify trend countries as well as steps taken to identify trend items and scales.

Does computer access equal skill or does practice make perfect?
Yemurai Tsokodayi
(American Institutes for Research, US)

Access to technology has boomed over the past 20 years. World Bank data shows that between 1995 and 2016, access to the internet has increased from less than 1 percent of the world's population to nearly 46 percent. During this time, technology has become very prevalent in schools as both a learning tool and as a literacy area. Despite the sharp rise of computer use in schools, research has shown mixed results on the benefit of computer access in schools. Studies using large-scale international education data have shown that at the country level, students in countries with higher levels of ICT development tend to have higher achievement, while at the student level there can either be minimal to no relationship, or positive effects on achievement, depending on the type of computer use. These mixed results highlight the continued need to understand the way technology is used in schools if students are to navigate online learning environments. Using the combined PIRLS and ePIRLS datasets, the study looks at the level of access and use students across 16 education systems have to computers and online information. Specifically, the study explores where fourth grade students learned how to use a computer and find information online, and the extent to which fourth grade students use computers in classrooms and at home. The study also examines the relationship of these variables to students' online reading achievement.

Analyzing ePIRLS results internationally after adjusting for students' offline reading ability
Martin Hooper
(American Institutes for Research, US)

Online reading has become the primary way that people acquire information, and it necessitates a skill set that goes beyond the skills needed for traditional offline reading. Recognizing this, the IEA's PIRLS assessment was extended for 2016 to include ePIRLS, an assessment of online informational reading. With the rapid integration of technology in schools and daily life, there has been increasing concern about the existence of a digital divide—the belief that advantaged students have more access to technology and use it more, and therefore thrive in the online environment. As such, many are concerned that these differences in access and use further increase the gap between advantaged (high socioeconomic status students) and disadvantaged (low socioeconomic status students) students. Exploiting the PIRLS/ePIRLS 2016 design, where the same students took both PIRLS and ePIRLS and therefore had reading score estimates on both assessments, this paper examines the evidence on whether students with more access to computers and who use them more do better in online reading (ePIRLS) after controlling for their offline reading ability (PIRLS). By conducting such analysis after controlling for students' offline reading, we can come to a better understanding of the evidence supporting the notion that differences in computer access and use exacerbate differences in online reading.

Analyzing IEA data in R: Introducing the EdSurvey package and its direct estimation function
Paul Bailey
(American Institutes for Research, US)

Analyzing large-scale assessment data requires substantial machinery to correctly handle weights and plausible values. The EdSurvey package for the R programing language allows users to quickly download IEA data, search for variables, use exploratory data analysis techniques, read in and modify data, and then run analysis that correctly accounts for the plausible values and survey weights. These methods are verified, for each survey, to agree with existing methods when they exist. Recently, EdSurvey added two new exciting modules, multiple imputation for missing data, and direct estimation. Using simulations, we explore how multiple imputation and then direct estimation performs relative to using the normal methods, which are known to be biased.

Clarifying TIMSS Advanced mathematics 2015 results: A didactical approach through levels of mathematical knowledge operation

Franck Salles
(DEPP, Ministry of Education, France)

TIMSS Advanced provides information on mathematics and physics achievement of students having completed scientific education at the end of their secondary schooling. The target of this study is Grade 12 scientific track students in France. IEA, which has been running the project since its birth in 1995, has chosen to assess students' skills as closely as possible to what is actually taught in their country. In mathematics, the assessment framework therefore describes a set of content and cognitive domains belonging to the intersection of the curricula of participating countries. 2015 TIMSS Advanced results for France, especially on the cognitive domain "applying," raise issues and incite us to analyze more finely the mathematical tasks within TIMSS Advanced mathematics items. Using the French sphere of didactics as theoretical background, we conduct an analysis of mathematics items, via our a priori task analysis tool based on levels of mathematical knowledge operation. It makes it possible to better account for all TIMSS Advanced questions, at a level of study (Grade 12 scientific track) that is rarely explored by international studies focused on primary (TIMSS) or the end of compulsory schooling (PISA). The analysis of all the items, under the eye of our didactic tool, illustrated by some examples, confirms the relevance of this model for describing standardized assessments' test material in mathematics. It also reveals the difficulty of designing a common framework for an international study that takes into account each participating country's curriculum, IEA's stated objective. At such an advanced level, TIMSS Advanced assesses the acquisition of taught knowledge depending on different conceptual constructs and national institutional choices in mathematics education.

Differential item functioning as a pedagogical tool

Jeppe Bundsgaard
(Aarhus Universitet, Denmark)

International large-scale assessments like the International Computer and Information Literacy Study (ICILS) provide important empirically based knowledge through the proficiency scales, of what characterizes tasks at different difficulty levels, and what that says about students at different ability levels. In international comparisons, one of the threats against validity is country differential item functioning (DIF), also called item-by-country interaction. DIF is a measure of how much harder or easier an item is for a respondent of a given group as compared to respondents from other groups of equal ability. If students from one country find a specific item much harder or easier than students from other countries, it can impair the comparison of countries. Therefore, great efforts are directed towards analyzing for DIF and removing or changing items that show DIF. From another angle, however, this phenomenon can be seen not only as a threat to validity, but also as an insight into what distinguishes students from different countries, and possibly their education, on a content level, providing even more pedagogically useful information. Therefore, in this paper, the data from ICILS 2013 is re-analyzed to address the research question: Which kinds of tasks do Danish, Norwegian, and German students find difficult and/or easy in comparison with students of equal ability from other countries participating in ICILS 2013? The analyses show that Norwegian and Danish students find items related to computer literacy easier than their peers from other countries. On the other hand, Danish and, to a certain degree, Norwegian students find items related to information literacy more difficult. Opposed to this, German students do not find computer literacy easier, but they do seem to be comparably good at designing and laying out, for example, posters and web pages. This paper shows that essential results can be identified by comparing the distribution of difficulties of items in international large-scale assessments. This is a more constructive approach to the challenge of DIF, but it does not eliminate the serious threat to the validity of the comparison of countries.

Nerds or polymaths? Performance profiles at the end of primary education

Olesya Gladushyna and Rolf Strietholt
(Technische Universität Dortmund, Germany)

The paper deals with the analysis of the data from combined TIMSS/PIRLS 2011 to explore the subject-specific strengths and weaknesses among fourth grade students worldwide. Previous research failed to take into account that students differ in terms of overall performance levels when investigating subject-specific strengths and weaknesses. Therefore, the present study uses transformed performances scores so that each student's performance score was centered by his or her overall performance. In contrast to previous research, this study identifies qualitatively different performance profiles in math, science, and reading. Cross-country differences in the distribution of the performance profiles and association with student characteristics validate the conducted analyses. Our findings suggest that a large proportion of students, who never use or sometimes use the language of test at home, are high achieving in mathematics but show low results in both science and reading. This could be explained by the peculiarities of the school curriculum that ensures a comprehensive training in math but requires extracurricular reading activities from students to master the reading skills. Moreover, the results demonstrate that boys are more likely than girls to achieve higher academic results in math, whereas girls are doing better in reading. Consequently, the applied approach in conducting the latent profile analysis represents a groundbreaking alternative to determining the qualitative differences in multidimensional proficiency model.

I know I can, but do I have the time? The role of teachers' self-efficacy and perceived time constraints in implementing cognitive-activation strategies
Nani Teig, Ronny Scherer, and Trude Nilsen
(University of Oslo, Norway)

Considerable research has demonstrated that teachers' self-efficacy plays a major role in implementing instructional practices in classroom lessons. Only few studies, however, have examined the interplay between how teachers' self-efficacy and the challenges that lie outside their control are related to their implementation of cognitive-activation strategies (CAS), especially in science classrooms. Using structural equation modeling and the Norwegian TIMSS 2015 data from Grades 4, 5, 8, and 9, we explored the extent to which teachers' self-efficacy in teaching science and the perceived time constraints explained variation in the enactment of general and inquiry-based CAS. Measurement invariance analyses showed that teachers had the same conceptual understanding of the measures across grades (metric invariance). Overall, the findings revealed that highly self-efficacious teachers reported more frequent implementation of both general and inquiry-based CAS, whereas teachers who perceived strong time constraints reported less frequent use of inquiry-based CAS. We further found that these relations differed significantly between primary and secondary school teachers, and discuss these findings in light of teachers' professional competences and the resources for science activities in primary and secondary education. We point to the theoretical implications of this study for enhancing the conceptual understanding of generic and specific aspects of CAS and the practical implications for teacher education, professional development, and educational policy as they reveal two potential aspects where teachers may need to be supported.

Session 5B: ICILS
Schooling and learning in the Digital Age: Evidence from ICILS 2013
Chair: Julian Fraillon (Australian Council for Educational Research, Australia)
Discussant: Eva Klemenčič (Educational Research Institute, Slovenia)
Room: A220

ICT-related school culture and differences in students' CIL: Results from IEA-ICILS 2013
Birgit Eickelmann[1], Julia Gerick[2], and Jeppe Bundsgaard[3]
(Paderborn University, Germany[1]; Universität Hamburg, Germany[2]; Aarhus Universitet, Denmark[3])

Aspects of school culture are attributed a special role in the development of schools and teaching. This is especially the case in the context of teaching and learning with ICT. The assessment of aspects of school culture is therefore a central component of international comparative large-scale assessments. In IEA's International Computer and Information Literacy Study (ICILS) 2013, it is displayed as an integral part of the study's theoretical framework. School culture can be allocated on the process level of school and as such is therefore considered relevant for students' competence in computer and information literacy (CIL). However, it is unclear to what extent the school culture is actually related to CIL. Comparing the level of different aspects of school culture between different education systems, differences become apparent. Based on the IEA ICILS 2013 data, this contribution therefore focuses on the relevance of ICT-related aspects of school culture in relation to students' CIL in six countries (the Czech Republic, Denmark, Germany, the Netherlands, Poland, and Switzerland). The findings show some significant differences in the average student CIL depending on the degree of school culture aspects in some educational systems. In addition to findings based on secondary analysis of ICILS 2013 data, expert statements from the aforementioned countries commenting and reflecting on the quantitative results are obtained in a small complementary study, which is also part of the presented research. The assessments of the experts underline that taking the different concepts of schooling and of teaching CIL lead to an in-depth view on school cultural aspects. Findings of this multi-perspective approach are discussed in relation to schools stakeholders' challenges of implementing an ICT-related school culture, and finally the challenges of measuring and comparing school culture as part of international large-scale assessments is discussed. Furthermore, some suggestions for future survey questions are given.

Explaining differences in students' computer and information literacy by analyzing response times in IEA-ICILS 2013
Corinna Massek, Kerstin Drossel, and Birgit Eickelmann
(Paderborn University, Germany)

Due to the increasing use of information and communication technology, computer-related skills are important for all students to participate in the digital age, with educational systems playing a key role. However, previous studies have shown differences in students' computer and information literacy (CIL). Although various approaches have been used to explain these differences, response times have never been taken into consideration. Based on data from the IEA study, ICILS 2013, of the Czech Republic, Denmark, and Germany, this secondary analysis examines to what extent response times can be used as an explanatory approach for differences in CIL also within different groups of students according to student background characteristics (gender, socioeconomic background, and immigrant status). At first two processing profiles using a latent profile analysis based on response times are determined—a fast processing profile and a slow processing profile. To reveal how these profiles are related to students' CIL also with regard to student background characteristics, descriptive statistics are used. The results show that in the Czech Republic and Germany students belonging to the fast profile on average have significant higher CIL than students of the slow profile. In Denmark, there are no significant differences. Concerning the students' background characteristics in the Czech Republic, a significant negative time-on-task effect occurs within all groups, except for the group of students with an immigrant status. In Denmark, no significant differences are found, whereas in Germany a significant negative time-on-task effect is determined for girls, students with a higher cultural capital, and students without an immigrant status. The results show that process data can be used to explain differences in students'

CIL. Further analysis should take into account other aspects of CIL (for example, computer-based self-efficacy or reading literacy).

Gender differences in computer and information literacy: Some findings from ICILS 2013
Sue Thomson, Eveline Gebhardt, Kylie Hillman, and John Ainley
(Australian Council for Educational Research, Australia)

Despite great advances over the 20th century in terms of gender equality for women, women are still underrepresented in many STEM fields, and the participation of young women in ICT subjects and careers has actually been declining over recent years. Reasons for this lack of engagement include a perception that ICT, as a male-dominated field, is not welcoming to women, less exposure to ICT at early childhood level, and even female students being more likely than male students to express a belief that advanced ICT subjects were boring. Yet, educational assessments in a range of countries have reported that, on average, female students scored higher than male students on measures of CIL. These results are intriguing because they run counter to the aforementioned commonly held expectations, and to the participation of women in ICT careers or further education. This report provides a systematic investigation of differences between male and female students in CIL, computer usage, and attitudes to computer technology, and also investigates the use of computer technologies in the classroom, the personal use of, and attitudes to, computer technologies of female and male teachers. These investigations are based on data from the IEA International Computer and Information Literacy Study (ICILS) conducted in 2013. Can these data provide insights into the apparent contradiction? Female students were found to perform relatively better on more general tasks while male students generally performed relatively better on more technical tasks, however differences were not large. There were few differences between female and male students' general ICT self-efficacy, however, male students had significantly higher specialized ICT self-efficacy than female students. Male students reported slightly higher levels of interest in, and enjoyment of, ICT than their female peers in almost all countries. In addition, interest and enjoyment appeared to be more strongly related to CIL among male students than among female students. Hence, the higher levels of CIL achieved by female students do not appear to be driven by interest and enjoyment of the area. The examination of gender differences in teachers' experiences, dispositions, and uses of ICT is that any differences are small and or not consistent across countries. Female and male teachers in secondary school do not appear to differ in the extent of their pedagogical use of ICT.

Session 5C: ICCS
Civics and citizenship education in countries around the world
Chair: Jens Bruun (Aarhus Universitet, Denmark)
Room: A405

Cross-national priorities for civic and citizenship education
Tim Friedman[1], Gabriella Agrusti[2], Valeria Damiani[2], and Jennifer Hong[1]
(Australian Council for Educational Research, Australia[1]; LUMSA University, Italy[2])

The ever-changing world we live in has led new challenges for countries in educating their youth to become citizens in the 21st century. This paper explores cross-national priorities for the implementation of civic and citizenship education utilizing data collected across twenty-four countries that participated in the International Civic and Citizenship Education Study (ICCS) 2016. Information was collected on the main aims and objectives of civic and citizenship education for students at the lower secondary level from educational policy documents as well as how these are perceived at the school level. Results suggest a degree in commonality across countries in the importance of key content areas of civic and citizenship education that are incorporated in curriculum, while also yielding some striking differences. Some differences in the conceptualization of civic and citizenship education and perceptions at the school level are explored.

Apathetic, distrustful, interested, passive, institutional, and participatory: Profiles of youth political participation in Latin America
Ernesto Treviño, Diego Alonso Carrasco Ogaz, Cristóbal Villalobos, and Natalia López
(Pontificia Universidad Católica de Chile, Chile)

Latin America presents a scenario characterized by high rates of disapproval in the face of political activity, political disaffection relative to the distance of the political process, and distrust of politicians, parties, and parliament. Multiple analysts have indicated that this malaise would be especially present in youth. Research has been conducted with emphasis on the relationship between the political-institutional system, youth participation, and the exercise of citizenship. It has been observed that although young people participate less in institutional channels, such as voting or participation in political parties, they develop other forms of civil and/or political participation, based on their participation in community actions. Most of the literature on forms of youth participation in Latin America has studied the dimensions of participation separately, using a "variable-centered approach." This research analyzes the different dimensions of youth participation simultaneously, using an "individual-centered approach," which seeks to configure subject profiles using different variables. This approach studies a set of variables to understand the similarities and differences between a specific group of subjects. It seeks to understand the willingness of young people to collaborate in different forms of political participation. Thus, the study contributes to the discussion on the groups or profiles of youth political participation that have been developed in the United States, Asia, and European countries, visualizing what "types" of young people exist with respect to the willingness to participate politically in this region. The research analyzes the connections and tensions between the different forms of participation in Latin America, as a way of contributing to the paradox between low institutional participation and increasing

non-formal participation. Finally, the present work analyzes how these profiles differ among countries, sociodemographic variables (sex, socioeconomic level, and schooling of parents), civic attitudes (support for gender equity, support for ethnic minorities, and support for authoritarianism) and school variables (school dependency, open discussion in classes, and student participation), to map the concentration and dispersion among the different typologies. The central research questions are: How many and what are the profiles of future youth political participation in Latin America? How do these profiles vary between countries?

Aiming for sense of citizenship across the Nordic countries: Comparing priorities in citizenship education 2009-2016

Idunn Seland[1], Lihong Huang[1], Jens Bruun[2], Cecilia Arensmeier[3], and Jan Markus Löfström[4]
(OsloMet-Oslo Metropolitan University, Norway[1]; Aarhus Universitet, Denmark[2]; Örebro University, Sweden[3]; University of Helsinki, Finland[4])

This study compares school principals' responses to the survey question "What are the three most important aims for citizenship education at your school?" in the International Civic and Citizenship Education Study (ICCS) focusing on Denmark, Finland, Norway, and Sweden both in 2009 and 2016. While lower-secondary students in these four countries are at the highest level in civic knowledge, they simultaneously report rather low interest in politics and civic activities. A common point of reference for the Nordic countries is the Nordic education model, aiming for social mobility, participation, and democratic citizenship. In this paper we ask whether principals' priorities for civic and citizenship education (CCE) point towards a Nordic model for such education, bridging the four national educational systems. We realize our comparison across the four countries through CCE curricular analyses and ICCS data. We find in our analysis a consistent and high support across the Nordic countries to promote students' critical and independent thinking, differing significantly from the international average. We discuss this finding reflecting on three different functions of education, where "critical thinking" is linked to the concept of subjectification, or the individual disposition to stand up against the existing social order. As stable democratic states rest on both critical and loyal citizens, they are faced with the dilemma of socializing their citizens to the democratic regime and at the same time promote a sense of critical thinking to uphold democratic principles and values. However, the Nordic model for CCE might be said to emphasize knowledge and critical thinking, rather than political engagement per se. When zooming in on the second and third priority for CCE in each country over time, the results are more varied. In this way, we may speak of one overarching Nordic ideal for CCE fanning out in four national varieties showing some internal similarities between countries and over time.

Italian teachers' preparation and training to teach civic and citizenship education topics and skills: Findings from ICCS 2016

Valeria Damiani[1], Elisa Caponera[2], and Bruno Losito[3]
(LUMSA University, Italy[1]; INVALSI, National Institute for the Educational Evaluation of Instruction and Training, Italy[2]; Roma Tre University, Italy[3])

Teachers' training and preparation has been acknowledged as crucial to the effective implementation of citizenship education. In Italy, in-service professional development activities are supported by national authorities but are often fragmented and the returns on teaching and learning processes are seldom monitored. The pre- and in-service trainings on civic and citizenship education (CCE) are of paramount importance in Italy since this learning area is not delivered as a separate subject in schools and is taught by non-specialized teachers. This contribution aims at delving into ICCS 2016 results and investigates whether Italian teachers' preparedness in CCE topics and their opportunities to attend trainings have an impact on Italian students' civic knowledge. In addition to this, it examines whether there are any differences between Italian Northern and Southern regions in relation to teachers' preparation and to their participation in professional development activities. Descriptive analysis and a path analysis with multilevel structural equation modelling were carried out. Results of the path analysis confirmed previous research on the relevance of students' SES on civic knowledge and showed that in schools where teachers have taken part in training activities on CCE, students participate more keenly in school activities. These findings corroborate the prominent role of school climate in developing students' participation (in schools that provide teachers with trainings on CCE, we found higher levels of students' involvement) and of pre-service professional development initiatives. Findings also highlighted differences related to Italian teachers' preparation and trainings, with this confirming, also for the learning area of CCE, the gaps between Northern and Southern Italy that characterize the Italian educational system.

Session 5D
Poster session
Chair: Seamus Hegarty (Publication and Editorial Committee, IEA)
Discussant: Lauren Musu (IEA)
Room: A403

Contributors to this session will give a 3 minute presentation of their poster. One poster will be awarded the Constantinos Papanastasiou Poster Prize. The poster prize winner will be announced during the closing ceremony. The poster award jury includes: Seamus Hegarty, Thierry Rocher, Dirk Hastedt, Christian Christrup Kjeldsen, and Elena Papanastasiou.

The Constantinos Papanastasiou Poster Prize:
A new conference initiative is the Constantinos Papanastasiou Poster Prize, in recognition of Professor Papanastasiou's enthusiastic long-term contributions to building and supporting the educational research community we enjoy today. Among his many accomplishments and achievements, Professor Papanastasiou was the original founder of the IEA International

Research Conference series, and hosted the inaugural event in Cyprus. Throughout his life he provided dedicated support and encouragement to young and emerging researchers.

Analysis of TIMSS 2015 framework: Input to the development of internationally-benchmarked 21st century Philippine assessment

Marilyn Ubina Balagtas[1], Danda Crimelda Buhain Garcia[2], and Dexter Cheng Ngo[2]
(Philippine Normal University, Philippines[1]; Rex Group of Companies, Inc.[2])

The Philippines implemented its K to 12 curriculum in 2013 for Filipino learners' holistic development and global competitiveness. One indicator of the curriculum's effectiveness is the country's performance in the Trends in International Mathematics and Science Study (TIMSS). To gauge Filipino students' readiness in such assessment, the alignment of TIMSS 2015 with Philippine's 2016 K to 12 curriculum in mathematics and science was examined. Through curriculum mapping, the study reveals gaps. The mathematics Grade 4 curriculum is more aligned with TIMSS 2015 than that of mathematics Grade 8, and science Grade 4 and 8. The results led to the development of assessment tools in Grade 4 and 8 mathematics and science that capture the competencies of the country's curriculum and those in the TIMSS Framework. The features and process in the development of these tools, benchmarked with TIMSS 2015, will be described. It is hoped that the results of this study could help explain possible outcomes of the Philippines' participation in TIMSS 2019 and inform further reform actions to make the K to 12 Program meet its goals.

An application of PRMSE to evaluate subscale score value in TIMSS 2015 fourth grade mathematics

Kondwani Kajera Mughogho
(University of Oslo, Norway)

TIMSS 2015 fourth grade mathematics test is comprised of three subscales: number, geometric shapes and measures, and data display. Demand for educational tests to provide subscale scores has increased as these scores have the potential to provide insights into specific content areas. Despite the potential benefits of subscale scores, they inherently possess lower reliability, may not provide distinctive information from each other, and may not provide information above and beyond that which is provided by the overall score. The study intends to empirically show the practical applicability and usefulness of proportional reduction in mean squared error (PRMSE) to evaluate subscale score quality in TIMSS. To do so, this study will (a) determine the general psychometric properties (i.e., observed score variance, subtest score reliability, and intercorrelation among subtest scores) of each subscale which may influence PRMSE on scores obtained from different subscale estimation methods, (b) determine the psychometric properties of each subscale for each participating country in order to examine the factor structure of each subscale in each participating country that may affect the PRMSE, and (c) analyze TIMSS 2015 fourth grade mathematics to assess the PRMSE values based on several item response theory based subscale estimation models (i.e., unidimensional and multidimensional) at population and subpopulation level. This paper will use response data from TIMSS 2015 fourth grade mathematics. The PRMSE will be applied to evaluate subscale scores, taken from scores drawn from item parameters that have been obtained from three subscale score estimation models. Data analysis will include: (a) an evaluation of the item parameter standard errors of measurement, (b) a model implied examination of the marginal reliability for each subscale, (c) a comparison of the subscale test information function (TIF) to those obtained for the overall score IRT, (d) model fit comparison, and (e) an examination of the estimated PRMSEs based on scores obtained from each model.

Modeling students' attitudes about learning mathematics in TIMSS 2015 context questionnaire

Militsa Georgieva Ivanova and Michalis P. Michaelides
(Univeristy of Cyprus, Cyprus)

Students' attitudes about learning, including students' motivation may have an impact on students' learning process and achievement. Therefore, it is important to use a valid measure of students' attitudes and motivations in learning mathematics in large scale testing programs, such as TIMSS. Teacher classroom behavior may also affect students' intrinsic motivation; therefore, it may have an effect on students' achievement. The present study aims to examine the structure of students' attitudes scale in the TIMSS 2015 student context questionnaires. All students' attitudes factors were considered to predict test performance. Data from TIMSS 2015 administration in the USA, including fourth and eighth grade students, was used for the present study. Results showed that the motivational construct in the TIMSS 2015 student context questionnaires can be described as a two-factor model for fourth grade students and a three-factor model for eighth grade students after negative wording effect is accounted for. Engaging teaching in math lessons is found to predict students' enjoyment in learning mathematics. Additionally, students' self-concept was found to predict test performance significantly, while students' attainment and utility values did not predict achievement on TIMSS 2015. An interesting finding was that enjoyment of math predicts test performance in that lower enjoyment is related to higher achievement. Overall, the presence of wording effect in TIMSS 2015 student background questionnaire was shown. In addition, teachers' classroom behavior predicted student intrinsic motivation, however only students' self-concept was found to significantly predict achievement. Therefore, teachers should concentrate on improving students' self-concept in order to enhance achievement in mathematics.

Perceived math self-efficacy as predictors of pursuing a STEM degree

Richard Carlos Limtuatco Velasco, Lee Kenneth Jones, and Yujiro Fujiwara
(Texas Tech University, US)

Numerous propositions and initiatives have fervently promoted the importance of infusing STEM (science, technology, engineering, and mathematics) curricula in the classroom with the hopes that newly graduated high school students would eventually pursue STEM careers. However, the undertaking of actually promoting STEM can be quite a daunting task, as pushing students through the STEM pipeline sees many of them leak through cracks and holes in the system; hence, an occurrence in the education world known as the "leaky STEM pipeline." Perhaps the greatest barrier in preventing students from pursuing a degree in STEM can be traced to their mathematics achievement and their belief in being able to perform well in mathematics. This study seeks to examine and analyze, in particular, high school advanced math students' self-efficacy beliefs and determine whether such beliefs play a role in predicting their choice of degree during their postsecondary studies. Findings reveal that male students who have high math self-efficacy and have parents with higher levels of education are most likely to pursue a STEM-related degree in postsecondary studies.

Achievement profiles in mathematics and science among Flemish fourth graders: An exploratory analysis
Lies Appels, Jerich Faddar, Sven De Maeyer, and Peter Van Petegem
(University of Antwerp, Belgium)

International large-scale assessment studies usually define *achievement profiles* as countries' achievement in specific subjects compared against international benchmarks. Moreover, previous studies investigated profiles within one specific subject or strived to label a country with one certain profile. An assumption that is supported by empirical results from this research is that learning in one subject develops at least partly in parallel with the learning in another. This suggests the idea of achievement as a multidimensional construct. The current study aims to describe *achievement profiles* as the recurring pattern of strengths and weaknesses in the three cognitive domains of the subjects mathematics and science within one country. In this way, a more detailed picture is drawn of the proficiency profiles within this country, in order to develop a deeper understanding of learners and their learning processes. The question thus arises whether we can distinguish groups of students with a different pattern in their performance within these domains or will it, in the end, come down to the idea of the multidimensional construct? In order to group similar students according to their achievement in mathematics and science, and thus examine student profiles, this study used the scores of Flemish fourth graders in TIMSS 2015. As the TIMSS data contain distinct performance measures for different mathematics and science domains, they are especially suitable for this research. The global k-means clustering method was applied. This method proves itself to be reliable when the actual number of clusters is not yet known. After a first trial, the top and bottom performers on both subject matters were excluded to elucidate the differences in the medium-performing group of students. Again, the method computed a six-cluster pattern. Findings revealed that the six profiles are significantly different in their mean achievement score on a subject, as well as in domain scores. However, the differences between domain scores within a cluster did not appear to be significant. Therefore, it was noted that a large majority of students did not score equally well on science and mathematics. However, their performance within a subject on the different domains seemed to be rather consistent.

Norwegian science teachers' content coverage in middle school: Patterns and factors
Stephan Daus
(University of Oslo, Norway)

National curricula specify the content to be covered by teachers in more or less detail, providing varying degrees of teacher autonomy. A teacher's decision-making in choosing which topics to cover in class and when to cover them is not just a process-factor, but also a neglected outcome variable as it indicates the pupils' opportunity to learn. This study examines content coverage patterns and the role of science teachers' knowledge and backgrounds for their content coverage decisions in Norway. Norway is characterized by high teacher autonomy, science education being taught as an integrated subject, and its participation in TIMSS 2015 with both grades eight and nine within the same schools. Responses from over 200 science teachers in each of the grades eight and nine within the same 142 schools are analyzed. Teachers' reported content coverage (as outcome) and preparedness in teaching on 22 science topics are modeled in generalized linear mixed models, where responses are cross-classified in teachers and topics. Other teacher variables included in the models are teachers' own domain specialization, and general pedagogical knowledge in teaching science. In the process, population-level descriptives of content coverage patterns across the schools are presented. Preliminary results indicate that there is more topic specialization (fewer topics covered) in ninth grade than in eighth grade, with much variation across teachers and topics in coverage for both grades. Clear grade-specific patterns across topics are identified, for instance, the TIMSS topic Light and Sound is generally not taught before grade ten. Teachers' preparedness in teaching content and general pedagogical knowledge in teaching science predict content coverage, and the two can compensate for each other, suggesting that teacher training can ensure opportunity to learn through increased content knowledge or increased general pedagogical knowledge. Teachers with only a bachelor degree and a chemistry background were found to teach more of chemistry topics than teachers with other backgrounds would cover of their respective domain specializations, indicating that chemistry teachers are biased in favoring their own subject. The intended curriculum in Norway is vaguely specified for three grades 8-10, and with little specifics on when to teach what. Thus, the findings offer specific information to textbook authors, educators, and newly trained teachers with little knowledge of what is currently being implemented in the classrooms. Moreover, the study might help teachers gain more specific knowledge from international large-scale assessments such as TIMSS, hence increasing their return of investment.

Citizenship through the IEA waves: Cross-cohort and country measurement invariance of two citizenship dimensions
Federico Viertel-Arze
(Universidad Santo Tomás, Chile)

In the last two decades, IEA has conducted three measurements that have gathered civic knowledge and attitudes towards citizenship (CIVED 1999, ICCS 2009/2016) in young students. The study of behaviors related to conventional citizenship (participation in political parties, voting in national elections) and related to social movements (participation in peaceful protests, protection of the natural environment) has been relevant. The same set of items on citizenship dimensions were included on the three waves of IEA studies. That allows the comparisions across cohorts in countries participating, provided that evidence can be attained that the measures in each study are equivalent. Multiple-group confirmatory factor analysis (MGCFA) is the most commonly used technique for measurement invariance/equivalence evaluation. MGCFA assumes equality of model parameters in all groups (cohorts and countries) and allows the evaluation of three hierarchical levels of measurement invariance trough the comparison of different models with increasing constraints. MGCFA was conducted between clusters of countries with cultural, linguistic, and historical similarities. Nordic: Norway, Denmark, Sweden, and Finland. Central and Eastern European: Bulgary, Estonia, Lithuania, Latvia, Rusian Federation, and Slovenia. South American: Chile and Colombia. Cross-Cohort and Country equivalence of measure on Conventional and Social-Movement-Related Citizenship dimensions allows latent mean scores comparison in three clusters of countries. Measurement model of Conventional Citizenship dimension was modified, excluding problematic indicators, with model variations across country clusters.

Measurement invariance in large-scale assessments using group-level fit: Patterns across scales and countries

Janine Buchholz[1] and Johan Braeken[2]
(German Institute for International Educational Research, Germany[1]; University of Oslo, Norway[2])

International large-scale assessments (ILSAs) aim at measuring and comparing cognitive and non-cognitive constructs between respondents from a large number of participating countries—an endeavor which requires measurement invariance (MI) across all participating countries to be established. In contrast to the cognitive part of these assessments, questionnaires are hardly subject to MI testing. This is problematic since most scientific publications are secondary analyses of non-cognitive constructs administered with the questionnaires. As a consequence, there is an operational need to investigate the appropriateness of comparisons of questionnaire constructs between countries participating in an ILSA. The most commonly employed method for MI testing is multigroup-CFA (MGCFA). Yet, in the presence of a large number of groups (i.e., countries participating in ILSAs), the method appears to be both practically infeasible and not informative for identifying groups of countries that can or that cannot be compared with each other. In the context of ILSAs, therefore, it is unclear what can be learnt or how to proceed when measurement invariance seems out of reach. In this study, we conducted a screening of MI across all scales and countries in the TIMSS 2015 and PISA 2015 questionnaires. Guided by recently developed indices of MGCFA group fit, we quantify the extent and variation of non-invariance issues, and we also report on identified patterns due to scale properties (e.g., length, response categories, previous use) and country characteristics (e.g., previous participation, geographic location, language groups, gross domestic product) that are consistent across the two ILSAs. These findings will help to identify country subsets for which meaningful comparisons are appropriate, and they may also be used to guide questionnaire development in the context of ILSAs in general.

Two decades of ILSAs: Meta-interpretation of student achievement results via comparison matrix

Mariusz Gałczyński
(McGill University, US)

Since the mid-1990s, ILSAs carried out by the IEA and the OECD have been administered in over half of all the world's countries. Widespread participation across perpetual testing cycles not only signifies tremendous investment of resources over more than two decades, it also represents fidelity to a global testing culture that has permeated all areas of education with its reinforcing nature. In the search for "best practices" among "top-performers," ILSA league tables have influenced stakeholders' perceptions of educational quality and have globalized education governance via policy reactions. In order to interpret ILSAs as validly and meaningfully as possible, the concept of ILSA literacy—an application of assessment literacy frameworks to globalization studies—demands that we frame results in broader context. Hence, this poster showcases the comparison matrix as a method of data visualization which reorganizes achievement scores in order to illustrate longitudinal trends and generate baseline references. Each matrix essentially serves as a meta-analysis of four major ILSAs (PISA, TIMSS, PIRLS, and ICCS)—though the term "meta-interpretation" more precisely describes the method of comparing analogous student populations within participating countries. With scores replaced by symbols denoting achievement in relation to scale averages, results are juxtaposed across multiple age/grade levels and in relation to multiple literacies. Rows group together relatively coincident ILSA administrations in order to identify the most comparable sets of achievement results, as sample populations are represented by contemporary student cohorts. The resulting comparison matrices facilitate easy visual identification of specific kinds of trends and incongruities. Three sample comparison matrices are depicted on the poster. The first matrix, constructed with data from countries participating in four major ILSA administrations between 2015-2016, helps viewers recognize high/low achievement trends as well as spot differences in achievement between STEM versus humanistic content areas, early versus later ages/grades, and IEA- versus OECD-sponsored ILSAs. The second matrix, constructed with data for the same time period, focuses on a single geographic region, subtracting countries from the Global North to prompt reflection over how the developing world participates in ILSAs. The third matrix, constructed with data from select countries participating in ILSAs since 1999, charts countries' histories of participation and achievement trends over time. As illustrations of ILSA literacy, these matrices elicit broader questions about the purpose of ILSAs, the legitimacy of policy transfer between systems, and the tangibility of goals inspired by "global" comparisons.

Science teaching practices and learning outcomes in high-performing Asian countries: Comparing TIMSS and PISA 2015
Anindito Aditomo
(Deutsches Institut für Internationale Pädagogische Forschung, Germany; University of Surabaya, Indonesia)

This study explores the extent to which TIMSS and PISA yield consistent patterns of associations between science teaching and affective as well as cognitive outcomes in 5 high-performing Asian countries. Items with similar content were identified and selected to represent two forms of instruction: "inquiry" and "interactive" science teaching. Multi-group ESEM of the measurement model established metric-level invariance across the 5 countries. Subsequently, multi-group multilevel SEM revealed markedly different patterns of teaching-learning relationships between the two assessments. The results show that PISA and TIMSS do not always converge or lead to consistent conclusions, especially regarding the associations between teaching and cognitive outcome (achievement). Results from PISA indicate that inquiry is associated with lower achievement, while interactive teaching with higher achievement. In contrast, the weak and non-significant relations found in TIMSS suggest that the relations between both forms of teaching and learning outcomes may be more contextual. A better understanding of the reasons behind these diverging findings is necessary before stronger conclusions can be made about the effects of interactive and inquiry-based science teaching. This paper argues that one of the more probable explanations is related to the source of data (student perceptions vs. teacher reports). That is, teachers and students may interpret descriptions of teaching practices (e.g. "students are asked to draw conclusions from experiments") in systematically different ways. Further research is needed to test this and other possible explanations.

School autonomy and the distribution of teaching quality: An international comparative perspective
Leah Natasha Glassow-Hill[1], Kajsa Yang Hansen[1], and Emilie Franck[2]
(University of Gothenburg, Sweden[1]; IEA[2])

This study investigates the relationship between school autonomy over teacher hiring and the distribution of teachers and teacher quality across schools. Evidence has shown that teacher quality is one of the key factors explaining the variation in the school achievement, as well as the socioeconomic gap in academic achievement. Often, more qualified teachers are concentrated in socioeconomically advantaged schools and such uneven distribution of teacher competences may be to some extent associated with country level policies in school autonomy over teacher hiring. While past research has shown that school autonomy in teacher hiring is associated with pedagogical segregation, a closer inspection of this relationship suggests this may not be the case. Hierarchical linear modeling (HLM) and hierarchical generalized linear model (HGLM) were applied to TIMSS 2015 data, focusing on school and country-levels simultaneously. Results show a general relationship between shared and full autonomy over shortage of teachers specialized in mathematics and science, but differential associations for cross-level interactions of school SES and autonomy. For average school level of teachers specialized in mathematics or mathematics education, there was no association with SES, but a moderating relationship with SES and level of autonomy.

Instructional sensitivity of the TIMSS science test: Adjacent-grades within schools
Stephan Daus[1], Agnes Stancel-Piątak[2], and Johan Braeken[1]
(University of Oslo, Norway[1]; IEA[2])

Evidence of instructional sensitivity of the TIMSS test has been unclear, possibly because its test development inadvertently omits items close to specific instruction, but also due to design limitations in previous instructional sensitivity analyses. We investigated the sensitivity of the TIMSS science test to Norwegian science instruction with a quasi-experimental design by analyzing responses from pupils in eighth and ninth grade classes within the same schools to 215 items, while accounting for school characteristics, content coverage, and a cohort effect representing cognitive development and general schooling. Preliminary evidence shows modest instructional sensitivity as, of all items, 16 percent show instructional sensitivity whereas 71 percent show sensitivity to cognitive development or schooling. Moreover, school-specific content coverage explains probability-correct to the same degree as the cohort effect.

Differences in determinants for ICT use and students' achievement in mathematics: Secondary analyses of TIMSS 2015
Julia Gerick[1], Birgit Eickelmann[2], and Kerstin Drossel[2]
(Universität Hamburg, Germany[1]; Paderborn University[2])

Based on the international IEA study TIMSS 2015, this contribution examines a secondary analysis of school-level determinants for the use of ICT in mathematics instruction in primary schools and their importance for student achievement in this competence area. In a comparison of six European education systems—Czech Republic, Denmark, Germany, Finland, the Netherlands, and Poland—we aim in a first step to identify school-level factors and combinations of factors which are essential for ICT use in instruction. Our goal thereby is to identify similarities between the countries as well as any country-specific hindering and supporting factors. In a second step, we examine the importance of ICT use for student achievement in mathematics. The results of the structural equation modeling show that most of the aspects related to ICT equipment, support and professional development are not linked directly to teachers' use of ICT in mathematics. Furthermore, no significant relation could be shown between the use of computers in mathematics instruction and the average student's

competency in this competence area. The results are discussed in the context of different traditions in the implementation of ICT in teaching and learning as well as in light of the methodological challenges of using large-scale assessment data.

Socioeconomic background and student achievement: TIMSS and PIRLS
Chair: Jan Mejding (Aarhus Universitet, Denmark)
Discussant: Plamen Mirazchiyski (Educational Research Institute, Slovenia)
Room: A403

Deconstruction of the negative social heritage? A search for variables confounding the simple relation between socioeconomic status and student achievement
Rune Muller Kristensen
(Danish School of Education, Denmark)

Negative social heritage was in the past often addressed through studies in which parents' educational level were linked to the educational level of their children. Today this phenomenon is addressed through a statistical relation between children's achievement and parents' socioeconomic status, the so-called ESCS measure (economic, social, and cultural status). The observed positive correlation between achievement and ESCS leads to a straightforward conclusion: Higher levels of socioeconomic status is associated with higher levels of attainment and vice versa with low values. The reason for calling this relation "negative social heritage," however, is due to an early sociological study in which only low status ("negative") families participated. The relation between achievement and ESCS has been known for years, but it was PISA 2003, by focusing on the numerical level of correlation ("weak" or "strong" negative social heritage), who brought considerable attention to the field in Denmark by comparing the values across countries. It was clear that Denmark ranked among the countries having the highest values of negative social heritage. The numerous ways of calculating an adequate estimate of ESCS will be discussed in the paper. The common reference for the ESCS measure is variables concerning parents' level of income, education, and occupational status. Consequently, an immediate interpretation of the level of correlation between student achievement and ESCS is that part of the student's level of attainment is due to the family background. The aim of the paper is to disclose what is in fact controlling this correlation, as it is clear that whatever definition you may choose for the calculation of ESCS, no one believes that the attainment level of the student can be derived operationally from background variables for the parents. TIMSS 2015 data has been the platform for exercising these analyses by involving all relevant student and parent variables, from "motivation" and "confidence" to "help to homework" and so on. The statistical analysis technique is log linear modeling, more specifically graphical modeling, analyzing higher order contingency tables (grouped variables) for main effects and interaction effects in the light of conditional independence graphs. With this, it is hoped to reveal variables that may act as confounders for the simple correlation between student attainment and ESCS. The results, however, point to a conclusion that none of the variables available can eliminate or diminish the basic correlation and one is consequently left with an open question: Are the variables determined by the TIMSS framework comprehensible for such analyses?

Effects of early tracking on performance and inequalities in achievement: Combined evidence from PIRLS, TIMSS, and PISA
Andrés Strello, Rolf Strietholt, Charlotte Siepmann, and Isa Steinmann
(Technische Universität Dortmund, Germany)

There is a recurrent discussion about the merits of the selective versus comprehensive school systems. The arguments about school placement policies, also called tracking, rest on a supposed tradeoff between equity and efficiency. The aim of the study was to determine the effect of early tracking in comparison to late tracking on the level of student achievement and two different measurements of educational inequality. The present research used a difference-in-differences approach and makes use of the pooled available data from PISA, PIRLS, and TIMSS studies with a total sample of more than two million students from 75 countries. The study indicates that early tracking increases both the dispersion of test scores and social inequality, and mixed effects on the mean performance of countries. Our study not only provides strong empirical evidence for effects of early tracking, but also illustrates the advantage of using the large pool of data available to address issues of stability and significance in international studies that suffer from the natural limitation of a low number of countries as the unit of analysis.

High achievement in mathematics and science: A multilevel analysis of TIMSS 2015 data for Ireland
Vasiliki Pitsia[1], Anastasios Karakolidis[1], and Gerry Shiel[2]
(Dublin City University, Ireland[1]; Educational Research Centre, Ireland[2])

High achievement at school has been linked to numerous benefits at individual and societal levels. It is important that education systems promote and reward high achievement, especially the knowledge and skills that are deemed necessary for living and working in the 21st century. However, research in the area of high achievement is scarce. This study examines student, home, class, and school factors associated with high achievement in mathematics and science among fourth grade students in Ireland using the Trends in International Mathematics and Science Study (TIMSS) 2015 data. Ireland provides a particularly interesting context because even though students have often performed well on national and international large-scale assessments of mathematics and science on average, there is a notable absence of high achievers and the scores of higher-achieving students have tended to be lower than their counterparts in other countries. Based on the results of multilevel logistic regression models of achievement, students' confidence, the resources for learning available in their

homes, and their views on engaging teaching in mathematics and science lessons were statistically significant predictors of high achievement in both subjects at the student level. In science, high achievement was also predicted by students' liking of science and significant gender differences were found in the distribution of high achievers. At class/school level, there were no statistically significant predictors of high achievement. The models explained a considerable proportion of the variance in mathematics and science high achievement. Implications for educational policy, practice, and future research are discussed.

Session 6D: Engaging with IEA data
Free online data platforms and tools for education researchers
Authors: Falk Brese[1], Yemurai Tsokodayi[2], and Yuqi Liao[2] (IEA[1]; American Institute for Research, US[2])
Room: A405

There are a number of data platforms and tools freely and publicly available on the web for supporting education researchers who wish to use existing data from international large-scale assessments (ILSAs) for a new research project. They vary considerably, and the primary objective of this session is to introduce participants to the platforms/tools and to show them when and how to effectively use these. Focusing on IEA studies, three platforms/tools will be presented, all of which have been developed with funding from the National Center for Education Statistics (NCES), US Department of Education:

- The ILSA Gateway is an excellent starting point for researchers, allowing them to gain both a first overview of the ILSAs that have been undertaken by the IEA and other organizations and to gain deep insights into the studies. The Gateway facilitates the development of research questions, the selection of the appropriate ILSA(s) to answer them, and access to resources on the external study websites.
- The International Data Explorer (IDE) supports researchers in their actual analyses of student and adult performance on ILSAs and of the related contextual factors, allowing them to run statistical tests and to generate customizable tables, charts, and maps. The IDE is of particular interest to all those wanting to use data from IEA studies (PIRLS and TIMSS) and OECD studies (PIAAC, PISA, and TALIS). With the data housed under one roof, the IDE includes data for all years from these ILSAs and for more than 100 education systems. The session will focus on IEA PIRLS data to demonstrate some of the available functions within the IDE.
- The Country Profiles and Comparisons (CPC) tool is of particular interest for researchers who want to examine and compare countries using assessment data gathered in ILSAs conducted by the IEA and the OECD and contextual data from other sources, including OECD's *Education at a Glance*, the World Bank Databank, and the CIA World Factbook.

Use of these tools will be demonstrated using example research questions. Afterwards, participants will have time to try out these tools by developing and working on research questions that could be addressed with available ILSA data. The authors will be supporting participants with using the tools, developing research questions, and analyzing ILSA data.

PRACTICAL INFORMATION

WIFI

If you are a student or employee at AU you can setup your wireless connection by using the "AU"-button. If your home institution uses Eduroam, you should be able to log on via the network: Eduroam.

Guest wifi
To use the guest wifi at DPU, please take the following steps:
1) Activate wifi on your device
2) Choose the 'AU-Guest' wifi
3) Open an internet browser
4) The browser will automatically redirect you to the login screen. (If not, please visit ww.au.dk). You need an account at one of the following to use the internet: Facebook, Google Drive, LinkedIn, SMS (Only Danish phone numbers), Microsoft-account (outlook.dk, hotmail.com.) After the initial configuration you will be logged on automatically in the future.

RECEPTION

The reception will be held in the Copenhagen City Hall
Address: Rådhuspladsen 1, 1599 København, Denmark
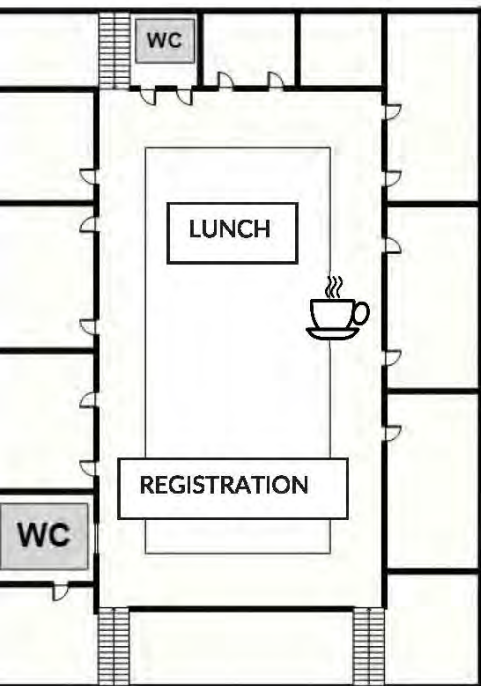18:00 – 19:30
Buses will be available to take guests to the event. The buses will leave at 17:20 from the front entrance to the conference building on Tuborgvej road.
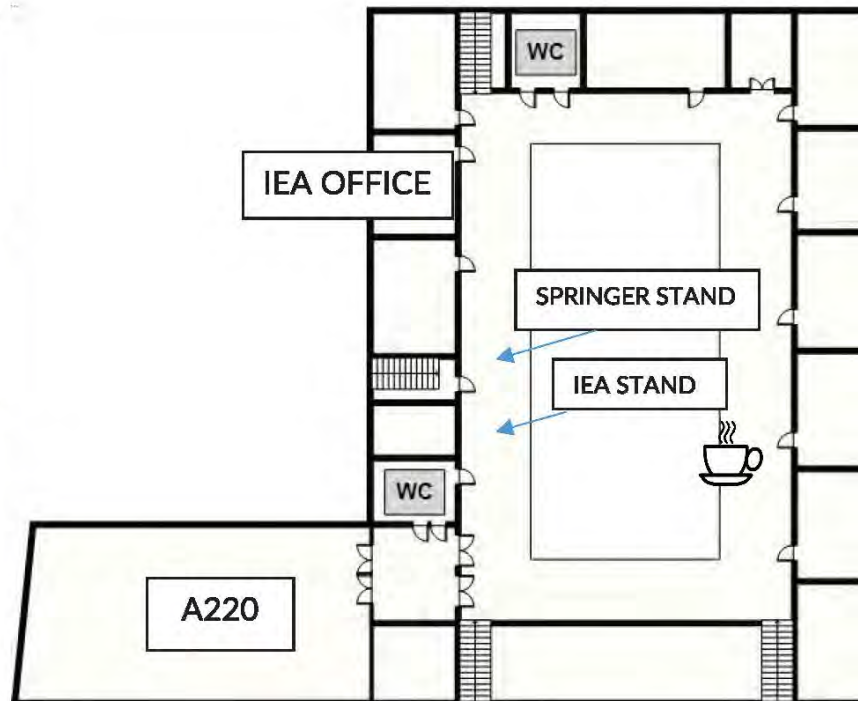
CERTIFICATES OF ATTENDANCE

If you would like a certificate of attendance, or have indicated to us already that you would like a certificate, please come to the registration desk on Friday (28 June) during the lunch break.
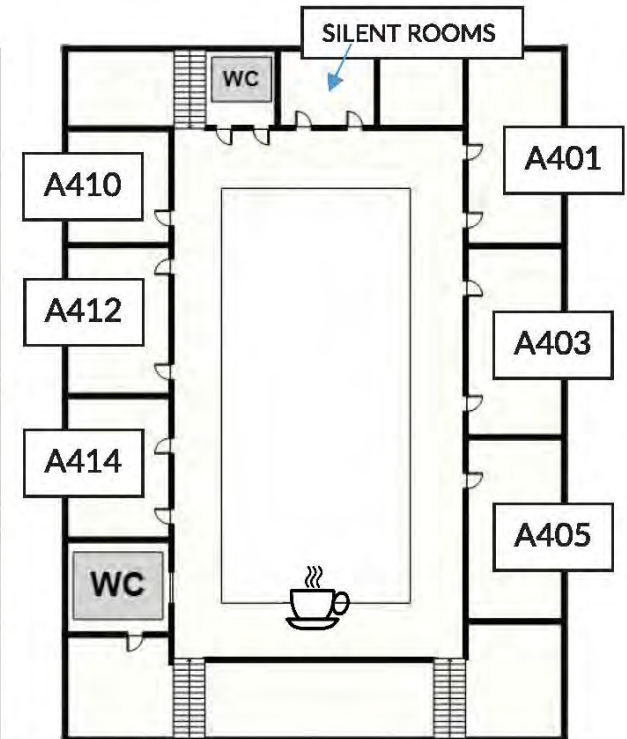
FLOOR PLANS

1ST FLOOR

WC

LUNCH

REGISTRATION

WC

2ND FLOOR

WC

IEA OFFICE

SPRINGER STAND

IEA STAND

WC

A220

4TH FLOOR

SILENT ROOMS

WC

A410

A412

A414

WC

A401

A403

A405

*Opening and closing ceremonies will take place in room A220

IEA encourages all presenters to consider adapting their papers for submission to

Large-scale Assessments in Education

Journal editors (Dirk Hastedt, Leslie Rutkowski, Sabine Meinck, and Eugenio Gonzalez) are available to answer your queries during this conference

Large-Scale Assessments in Education
Improving the science of large-scale assessment
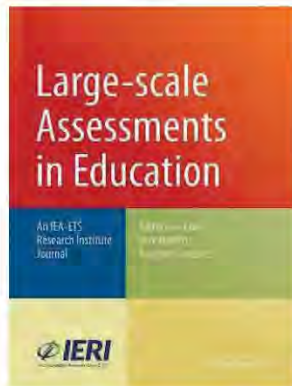
This journal is a joint publication of the IEA-ETS Research Institute (IERI). Articles contribute to the science of large-scale assessments, help disseminate state-of-the-art information about empirical research using these databases and make the results available to policymakers and researchers around the world.

Articles suitable for publication in the journal make use of data collected by programs, such as: IEA's TIMSS, PIRLS, ICCS, ICILS, and SITES; US-NAEP; OECD's PISA, PIAAC, and TALIS; IALS, ALL, and others. The journal also accepts Software Articles that showcase tools for educational assessment and research.

If you have questions or concerns about whether your article adheres to the purpose of this journal, please visit the journal website:
https://largescaleassessmentsineducation.springeropen.com/
Or send an informal email to one of the Editors.

Large-scale Assessments in Education operates a double-blind peer-review system. Submitted manuscripts are always reviewed by two or more experts, who will be asked to evaluate whether the manuscript is scientifically sound and coherent, whether it duplicates already published work, and whether or not the manuscript is sufficiently clear for publication.

The publication costs for *Large-scale Assessments in Education* are covered by the IERI Research Institute, thus authors do not need to pay an article-processing charge. Copyright on any open access article in a journal published by SpringerOpen is retained by the author(s).

Articles are widely promoted through social media, email updates, table of contents, email alerts, postings on the SpringerOpen homepage, and press releases to the general and scientific press—all resulting in increasing levels of access for each article. SpringerOpen journals are also promoted at many major scientific conferences to bring your work to the attention of professionals in your field.

We look forward to your submission!

SPRINGER NATURE

Springer Nature is one of the world's leading global research, educational, and professional publishers, home to an array of respected and trusted brands providing quality content through a range of innovative products and services. Springer Nature is the world's largest academic book publisher, publisher of the world's most influential journals, and a pioneer in the field of open research. The company numbers almost 13,000 staff in over 50 countries. Springer Nature was formed in 2015 through the merger of Nature Publishing Group, Palgrave Macmillan, Macmillan Education, and Springer Science+Business Media.

In addition to the *Large-scale Assessments in Education* journal, Springer publishes the open access *IEA Research for Education Series*. Find out more: www.springernature.com or visit the **Springer information stand.**

THANK YOU

IEA and the Danish School of Education, Aarhus University, would like to thank all participants for joining in the 8th IEA International Research Conference (IRC). The IRC was founded in 2004 to foster collaboration and knowledgesharing of findings related to IEA studies. The IEA IRC provides an international forum for those working with IEA study findings and data to exchange ideas and information on critical educational research issues in a comparative and global context. We hope you have found the conference informative and valuable. We would like to welcome you to please join us for the next IRC in the United Arab Emirates, 2021.

ABOUT THE 9TH IRC

23–25 JUNE 2021, UAE
The 9th IEA International Research Conference will be hosted by the UAE Ministry of Education (Pre-conference workshops will be held from the 21-22 June, 2021)

A WELCOME FROM THE UAE MINISTRY OF EDUCATION

Since the creation of the United Arab Emirates, education has been the foundation on which we have built our progressive and forward-thinking society. The UAE has been dedicated to creating pioneering education institutions characterized by high-quality learning environments with access to modern technology that enable generations of students to complete their education. The Ministry of Education has been working hard to develop curriculum learning standards and national assessments that reflect best global practices as well as promoting a culture of creativity and innovation. As an IEA member and participant in TIMSS and PIRLS, the UAE is proud to collaborate with and support IEA as hosts of the 9th International Research Conference.

WWW.IEA.NL

IRC 2019 | IEA | UNIVERSITAS ARHUSIENSIS · SOLIDUM PETIT IN PROFUNDIS | DANISH SCHOOL OF EDUCATION
AARHUS UNIVERSITY

IEA

IRC 2019

This is to certify that

Anindito Aditomo

participated in the

8th IEA International Research Conference

26-28 June 2019
Danish School of Education, Aarhus University
Copenhagen, Denmark

Dirk Hastedt
IEA Executive Director

Christian Christrup Kjeldsen
Deputy Head of School (DPU) and
Director for Research