

Session Wed 29, 10:00 - 11:00 1

29 August 2018 10:00 - 11:00

Auditorium

SIG 18 - 23 Keynote Session

SIG 18 - 23 Keynote Session "Intelligent Inspection – Towards research-informed practice"

Keywords: Accountability and school inspections, Curriculum change and development, Educational policy, Improvement

Interest group:

Chairperson: Roel J. Bosker, Rijksuniversiteit Groningen, Dept of Education and GION, Netherlands

Intelligent Inspection – Towards research-informed practice

Keywords: Accountability and school inspections, Curriculum change and development, Educational policy, Improvement

Presenting Author: Daniel Muijs, University of Southampton, United Kingdom

Inspection is a part of many accountability systems across Europe, though the extent to which inspections are high stakes differs (Ehren et al, 2013). However, both the impact of inspection in terms of school improvement and its reliability and validity have been questioned (Altricher & Kemethofer, 2015; Gaertner et al, 2014). In terms of impact, we can draw on a number of useful models, such as Ehren et al (2015) which posit the impact of inspection as not necessarily resulting directly from inspection leading to change in the inspected school, but as much or more in the ways they frame behaviours by setting up parameters which define good practise. In terms of reliability and validity, some inspection systems internationally have attempted to increase reliability by drawing up detailed frameworks and minimising judgement, while others have continued to emphasise the primacy of judgement (Richards, 2001; Dederling, 2015; Earley et al, 2017). While it would seem at face value that the first approach is the one to be preferred if we are looking to develop a reliable system in high stakes contexts, this can easily lead to unintended consequences such as the reification of the framework and a box ticking approach to accountability that can be detrimental to developing a thoughtful and self-improving system, which we would see as a key aim of an intelligent accountability system. Such a system needs to be self-evaluative and corrective, which suggests that research has a key role in its development and maintenance. In this presentation I will describe a conceptual model for an intelligent inspection system, and illustrate the role of research using the examples of two major research projects focussing on key elements of inspection in England, curriculum and classroom observation. We will discuss both how as an inspectorate we can conduct research, and what the issues are, and how that research can then inform practise.

Session Wed 29, 11:30 - 13:00 1

29 August 2018 11:30 - 13:00

Conference room 2

Single Paper

Assessment and Evaluation

Single Paper Secondary education

Keywords: Class and school composition, Effectiveness and efficiency, Improvement, Quasi-experimental research, Secondary data analysis, Secondary education, Social aspects of learning and teaching, Student background

Interest group: SIG 18 - Educational Effectiveness

Chairperson: Christine Rubie-Davies, University of Auckland, New Zealand

Studying abroad during high school: Determinants and effects

Keywords: Class and school composition, Social aspects of learning and teaching, Student background, Secondary data analysis

Presenting Author: Nicolas Hübner, University of Tübingen, Germany; **Co-Author:** Ulrich Trautwein, University of Tübingen, Germany; **Co-Author:** Benjamin Nagengast, Eberhard Karls Universität Tübingen, Germany

Advanced skills in foreign languages such as English—the “lingua franca” of our time—count among the core educational objectives of secondary school. In a globalized world, extended stays abroad during school education are considered an important opportunity for acquiring transnational human capital. However, little is known about the predictors for studying abroad during school education and about the effects of such stays. Using a large longitudinal data set, considering population data of students from Hamburg (Germany), this study tries to investigate the determinants and effects of studying abroad during school time. For this purpose, three waves of the LAU longitudinal study (Learning Aspects of the initial situation and Learning Development) were considered. The total sample consisted of $N = 5,566$ students. Of these, $n = 717$ students studied in a foreign country during Grade 10-13 for a minimum of 6 months ($n = 530$ studied in an English-speaking destination). Hypothesis and procedures for addressing the research questions regarding the determinants and effects of going abroad during school time were pre-registered at the open science framework (www.osf.io). First results of multiple logistic regression models are in line with previous findings by Gerhards and Hans (2013) and indicate that especially socioeconomic aspects and achievement are related to the decision to study abroad. Most interestingly, school SES also positively predicted going abroad over and above the individual socioeconomic background and private vs. non-

private school indicators. Final results on determinants and effects as well as implications for policy and practice will be discussed.

Impact and sustainability of the effect of DASI on student achievement gains in secondary education

Keywords: Effectiveness and efficiency, Improvement, Secondary education, Quasi-experimental research

Presenting Author: Yiannis Ioannou, University of Cyprus, Cyprus; **Co-Author:** Leonidas Kyriakides, University of Cyprus, Cyprus; **Co-Author:** Anastasia Panayiotou, University of Cyprus, Cyprus

This paper advocates for the use of the Dynamic Approach to School Improvement (DASI) to promote quality in education. A number of empirical studies as well as two meta-analyses supported the validity of the theoretical framework of DASI. All studies investigating the impact of DASI on school improvement were focused on primary education. This study aims to examine the impact that DASI can have on school effectiveness when being implemented in secondary education and the sustainability of its results one year after its completion. This quasi-experimental study was conducted during three school years. DASI was offered to a sample of six lower secondary schools which participated on a voluntarily basis while another sample of six schools with similar characteristics comprised the control group. To measure the impact of DASI on student achievement, tests in mathematics were administered to all students of the school sample (n=3701) at the beginning and end of each school year. Teachers of the experimental schools completed a questionnaire measuring the school factors of the dynamic model at the beginning of each school year. Multilevel analysis revealed that schools in the experimental group managed to achieve better results compared to those in the control group at the end of each school year of the project, while students' achievement gains were bigger for schools that have continuously implemented the program over the three-year period. Direct as well as indirect effects of school policies on student achievement gains were also identified. Implications of findings are drawn.

Session Wed 29, 11:30 - 13:00 2

29 August 2018 11:30 - 13:00

Leekstermeerzaal

Roundtable

Educational Policy and Systems

Roundtable Data based decision making

Keywords: Attitudes and beliefs, Data-based decision making and data use, Improvement, Primary education

Interest group: SIG 23 - Educational Evaluation, Accountability and School Improvement

Chairperson: Tina Montreuil, McGill University, Canada

Data-based decision making in primary education : assessing stakeholders' attitude

Keywords: Attitudes and beliefs, Data-based decision making and data use, Improvement, Primary education

Presenting Author: Lionel Alvarez, HEP-Fribourg | Université de Fribourg, Switzerland; **Co-Author:** Pierre-François Coen, Applied University for Teacher Education – Fribourg, Switzerland

Data-based decision making (DBDM) is more and more promoted in primary education, but it requires practice changes from every stakeholder. Assessing school readiness for DBDM before its implementation can be crucial for its broad adoption and proper use, because attitudes and skills influence how DBDM can be accepted and operated. Such an assessment can open up to a personalized DBDM implementation process, based on the school readiness results. A Delphi technique starting with several statements about DBDM interests, challenges, and impacts, is proposed to assess the school readiness for DBDM. Through the iterative process integrated within the Delphi technique, school stakeholders will come to a consensual perception of DBDM implementation needs and challenges, as well as a variety of individual perceptions. This round table session is organized to discuss this process of school readiness assessment with the Delphi technique. Attendants will be asked to judge the research protocol thanks to their knowledge about DBDM implementation, and thanks to their expertise about the Delphi technique.

Session Wed 29, 11:30 - 13:00 3

29 August 2018 11:30 - 13:00

Conference room 1

Single Paper

Educational Policy and Systems, Learning and Instructional Technology

Single Paper Accountability & networks

Keywords: Accountability and school inspections, Case studies, Comparative studies, Educational policy, Improvement, Learning communities, Mixed-method research, Qualitative methods, Quantitative methods, Quasi-experimental research, Teacher professional development

Interest group: SIG 18 - Educational Effectiveness, SIG 23 - Educational Evaluation, Accountability and School Improvement

Chairperson: Sirkku Kupiainen, University of Helsinki, Finland

Accountability of educational networks: findings from four case studies

Keywords: Accountability and school inspections, Educational policy, Case studies, Comparative studies

Presenting Author:Melanie Ehren, UCL Institute of Education, United Kingdom; **Co-Author:**Joe O'Hara, Dublin City University, Ireland; **Co-Author:**Martin Brown, DCU, Ireland; **Co-Author:**Jorgen Svensson, Universiteit Twente, Netherlands; **Co-Author:**Rossitsa Simeonova, Sofia University, Bulgaria; **Co-Author:**Gerry McNamara, Dublin City University, Ireland; **Co-Author:**Yonka Parvanova, Sofia University, Bulgaria

This paper examines the accountability of educational networks through inspections and seeks to answer the question *How are educational networks inspected, and what is the impact of these inspections on the functioning and performance of networks?* Four cases of European Inspectorates of Education -England, the Netherlands, Bulgaria, and Northern Ireland – are presented. Each of these has reconsidered their roles and ways of working in the context of network governance in recent years. The paper provides a conceptual framework building on existing research relating to networks governance and accountability and offers a detailed snapshot of each of the cases using this framework. The conclusions from this study argue that while holding networks to account can have positive impacts, there are also unintended side effects. The study suggests that external accountability of networks seems to be most effective when 1) single member and network-level accountability are well aligned, 2) external inspection and monitoring is responsive to the purpose, properties and the development of the network, 3) and focuses on the performance of networks (instead of compliance to regulation), and 4) where there is high trust between schools, network, inspector/Inspectorate (and active management of that trust).

Professional Learning Communities and Innovations in education - an implementation study

Keywords: Improvement, Learning communities, Teacher professional development, Quasi-experimental research

Presenting Author:Catalina Lomos, LISER - Luxembourg Institute for Socio Economic Research, Luxembourg

Policymakers strive for success in education and recommend implementing educational reforms and innovations, which are increasingly necessary in a culturally and socially changing environment. However, changes in education can be demanding on schools, and teachers require specific support structures and mechanisms to help motivate them to change. Previous research has indicated that teacher commitment is improved and higher student outcomes are achieved when teachers work in Professional Learning Communities (PLCs) within school organizations (Lomos, 2012). Following an experimental design, the main study aims at organizing schools as Professional Learning Communities during the implementation of the MathemaTIC innovation process. In order to insure the feasibility of this large planned project, the "intervention" (the PLC training program), is being developed and it is under implementation starting January 2018. The aim of this implementation study is to monitor this new program and take note of the process. This implementation study will allow us to set-up a highly performing intervention and design a complementary study of intervention fidelity during the main intervention planned (Mendive, Weiland, Yoshikawa, & Snow, 2016).

Considerations about public interests: A school board perspective

Keywords: Educational policy, Mixed-method research, Qualitative methods, Quantitative methods

Presenting Author:Marinda Spithoff, Rijksuniversiteit Groningen, Netherlands; **Co-Author:**Gerry Reezigt, Rijksuniversiteit Groningen, Dept of Education and GION, Netherlands

The main responsibilities for Dutch school boards are assuring the quality of education, the accessibility of education and the efficiency of education. School boards need to find balance in these public interests. This study examines which considerations school boards make for their decision-making, when their schools face difficulties such as decreasing student numbers. This study used a mixed methods approach with qualitative as well as quantitative data. A sample of Dutch school boards (n=34) that have experienced, or will experience a decrease in student population, participated. Through a questionnaire and an interview, the chairs of the school boards were questioned about their experiences and decisions, and there were asked to respond to a case of a school board facing difficulties related to a decreasing student number. The finding from the interviews were compared with data from the questionnaire and data about context characteristics of school boards, to determine which characteristics affect the considerations of school boards. The chairs of the school board often mentioned the importance of taking into account the quality of education in their decision-making. However, when school board directors decided to intervene on schools that were experiencing a decreasing student number, this decision was mainly based on efficiency arguments. The accessibility was also mentioned often. In some school boards there seems to be a difference between the advice school chairs give and their own practice.

Session Wed 29, 11:30 - 13:00 4

29 August 2018 11:30 - 13:00

Auditorium

Symposium

Assessment and Evaluation

Smartphone assisted student feedback on teaching quality put to the test

Keywords: Effective instruction and teaching, Effectiveness and efficiency, Experimental studies, Psychometrics, Secondary education

Interest group:

Chairperson: Gerry Reezigt, Rijksuniversiteit Groningen, Dept of Education and GION, Netherlands

Discussant: Emmeline van der Scheer, University of Groningen, Netherlands

Teaching quality can be measured in several ways. Lesson observations are quite common, however, measuring the perceptions of the target group, the students, about teaching quality is not done very frequently yet. Modern technology enables the efficient collection and processing of student perceptions of teaching quality as well as feeding back the results to teachers in order to support the improvement of teaching quality. The digital Impact! tool was developed to provide teachers with feedback about the degree to which the lesson that just ended met 15 characteristics of effective teaching. In this symposium, the results will be reported of three studies into the Impact! tool. First, we will present the effects of the use of the tool by secondary school mathematics teachers on their teaching quality. This was investigated in a randomized controlled trial. Next, the results of a generalization study (G-study) will be reported, in which the total variance in teachers' teaching quality scores was decomposed into various variance components (the differences between teachers, measurements, students, items, and their interactions). In addition, three other indicators of psychometric quality of the Impact! tool were calculated. In the third study, we investigated the concurrent validity of the Impact! tool. This was done by analyzing to what extent the ratings of a lesson by students, by external observers and by the teacher himself/herself are similar, or differ from each other.

Does smartphone-assisted student feedback impact teaching quality?

Presenting Author: Adrie Visscher, Univ. of Twente, Unknown; **Co-Author:** Hannah Bijlsma, Fac. of Behavioural Sciences, University of Twente, Netherlands; **Co-Author:** Marjoleine Dobbelaer, University of Twente, Netherlands

In this study, student perceptions of teaching quality were measured by means of a smartphone application for providing teachers with feedback on their lesson. It was investigated if student feedback promoted teachers' insight into where they could improve their lessons, and promoted their reflection on their lessons. It was also investigated in what ways teachers worked on improvement, and whether the student feedback affected the quality of their teaching. Teachers reported that they gained insight into their improvements based on the student feedback. They did not seem to reflect statistically significant more on their lessons. Teachers reported improvement-oriented actions in response to the student feedback. According to students, teachers first slightly improved their teaching quality. However, the improvement did not sustain. Explanations for the findings are discussed and suggestions for future research are presented.

Psychometric quality of the Impact! tool

Presenting Author: Hannah Bijlsma, Fac. of Behavioural Sciences, University of Twente, Netherlands; **Co-Author:** Adrie Visscher, Univ. of Twente, Unknown

To obtain student perceptions of teaching quality efficiently, the Impact! tool was developed and the effect of using student feedback on teachers' teaching quality was investigated (as will be presented in the first presentation of this symposium). In this study, the psychometric quality of the Impact!-tool was investigated by developing an IRT-model on the students' scores. Reliability was established by conducting a generalization study on this IRT-model, which explained the variance of student perceptions of teaching quality based on differences between teachers, students, measurement moments, items, and the interactions between these factors. Results show that student perceptions are reliable measurements of teaching quality. Moreover, all items contributed in a high extent to the estimation of teaching quality, which is an indicator of construct validity.

To what extent do students, teachers, and external observers agree on lesson quality?

Presenting Author: Marjoleine Dobbelaer, University of Twente, Netherlands; **Co-Author:** Adrie Visscher, Univ. of Twente, Unknown; **Co-Author:** Hannah Bijlsma, Fac. of Behavioural Sciences, University of Twente, Netherlands

In this study, the concurrent validity of three different judgements of teaching quality was explored: that of external observers, teachers and students. Little is known about whether these judgements are similar. For 25 teachers, 3 lessons were rated by students using the Impact! tool at the end of the lesson. Teachers also rated each lesson, using a questionnaire with the same (reformulated) items as in the Impact! tool. All lessons were video-taped as well and rated by 3 trained external observers using an observation instrument with the same (reformulated) items as in the Impact! tool and the ICALT observation instrument. The results show to what extent the ratings are similar.

Session Wed 29, 14:00 - 15:30 1

29 August 2018 14:00 - 15:30

Conference room 2

Single Paper

Assessment and Evaluation

Single Paper Ratings of teaching quality 1

Keywords: Accountability and school inspections, Classroom climate, Educational policy, Effective instruction and teaching, Large scale assessment/large scale survey, Primary education, Psychometrics, Survey Research, Teacher professional development

Interest group: SIG 23 - Educational Evaluation, Accountability and School Improvement

Chairperson: Catalina Lomos, Luxembourg

The association between student survey and classroom observation measures of teaching effectiveness

Keywords: Educational policy, Effective instruction and teaching, Psychometrics, Survey Research

Presenting Author:Rikkert van der Lans, University of Groningen, Netherlands

This study examines the association between two measures of teaching effectiveness—a student survey and a classroom observation measure—to determine (1) whether the correlation is dependent on the number of classroom observations and (2) whether the correlation is dependent on whether the survey design is nested or partially nested. The sample includes 57 teachers who received three lesson visits and administered one student survey. Using generalizability theory, this study reveals that the correlation between the survey and observation measures increases as the number of lesson visits by different classroom observers increases. In all investigated scenarios, the correlation doubles when three lessons are visited by different observers instead of one. The results also indicate that correlations tend to be positively biased if the survey design is completely nested. Implications for both the evaluation and research practice are discussed.

Student ratings of teaching quality: Do they differentiate between different subjects?

Keywords: Classroom climate, Effective instruction and teaching, Primary education, Large scale assessment/large scale survey

Presenting Author:Svenja Rieser, University of Wuppertal, School of Education, Germany; **Co-Author:**Alexander Naumann, German Institute for International Educational Research (DIPF), Germany; **Co-Author:**Ruven Stahns, University of Duisburg-Essen, Germany; **Co-Author:**Heike Wendt, Institute for School Development Research (IFS), TU Dortmund, Germany

Student ratings of teaching quality are frequently referred to as a source of information on classroom processes. However, their validity is often doubted. Our study adds to the literature on the validity of student ratings by investigating whether primary school students' ratings of three dimensions of instructional quality (cognitive activation, classroom management, supportive climate) differentiate between different subjects. Drawing on data from the TIMSS 2015 student questionnaire we specified two multi-level CFA models (one for Mathematics, one for Science classes) with three factors on each level to check whether the data fitted the three dimensions assumed in the theoretical framework. We then specified a six factor model (three dimensions for each subject) for classes taught by the same teacher and one for classes taught by different teachers to analyse the correlations between the latent factors across subjects. The three-factor model showed an adequate fit to the data. When taught by the same teacher latent correlations between the factors across subjects were very high. When taught by different teachers the correlations were much smaller and only one of three correlations became statistically meaningful. Our results show that primary school students differentiate between different subjects and teachers in their judgements on teaching quality. This can be interpreted as an indication for the validity of student ratings.

Validity and reliability of student perceptions of teaching quality in primary education

Keywords: Accountability and school inspections, Teacher professional development, Primary education, Psychometrics

Presenting Author:Emmeline van der Scheer, University of Groningen, Netherlands; **Co-Author:**Hannah Bijlsma, Fac. of Behavioural Sciences, University of Twente, Netherlands; **Co-Author:**Cees Glas, Fac. of Behavioural Sciences, University of Twente, Netherlands

Lesson observations by external observers are often used to measure teacher quality. Another way is to use student perceptions of teaching quality. However, in primary education these perceptions are not used a lot due to doubts about the reliability and validity of perceptions of young students. In this study, a Bayesian IRT-model approach was used to investigate the validity and reliability of student perceptions of teaching quality. Furthermore, the student perceptions were compared with ratings of teaching quality by external observers. Three lessons of 31 teachers were recorded and rated by four raters. Grade 4 students (N=585) filled out a questionnaire that was used to measure their opinion about the lessons of their teachers. The analyses showed that student perceptions were valid and the student perception scales were reliable, although, not all items contributed to the same extent to the scales. Student ratings and lesson observations scores generally correlated moderately (ranging from $r = .18$ to $r = .50$). Higher correlations were found for scales with a similar content. Explanations for the findings and suggestions for future research are given.

Session Wed 29, 14:00 - 15:30 2

29 August 2018 14:00 - 15:30

Leekstermeerzaal

Single Paper

Assessment and Evaluation, Teaching and Teacher Education

Single Paper Interactions 1

Keywords: Accountability and school inspections, Achievement, Classroom climate, Effective instruction and teaching, Improvement, Longitudinal analyses/research, Primary education, Psychometrics, Quantitative methods, Teacher professional development

Interest group: SIG 18 - Educational Effectiveness, SIG 23 - Educational Evaluation, Accountability and School Improvement

Chairperson: Charalambos Charalambous, University of Cyprus, Cyprus

A systematic review of the qualities of classroom observation systems

Keywords: Accountability and school inspections, Improvement, Teacher professional development, Primary education
Presenting Author:Marjoleine Dobbelaer, University of Twente, Netherlands; **Co-Author:**Adrie Visscher, Univ. of Twente, Unknown

Teaching quality is often measured by means of classroom observation using a classroom observation system (COS). Generating valid and reliable scores when using a COS is not self-explanatory. Many authors point to issues that need to be taken into account when developing, selecting and/or using a COS in order to generate valid and reliable scores. We first developed a framework that brings together these issues and that can be used for evaluating COSs. The framework draws from three strands of literature: the literature on COSs, the literature on testing and performance assessment, and the literature on the argument-based approach to validity. We then evaluated 28 COSs for measuring teaching quality in primary education, using this evaluation framework. The COSs in this review were found by means of an extensive literature search. Over 20.000 publications were scanned and about 200 COSs were compared with our inclusion and exclusion criteria. The evaluation of the 28 COSs shows which teaching quality constructs the COSs aim to measure, and the theoretical basis for this definition. We will also report to what extent the COSs include scoring rules, rater training, require rater certification, and other aspects that can be included in a COS. We will present the research evidence with respect to the reliability of the COSs and examine the evidence available for justifying ratings of teaching quality and the use of these scores. This review provides insight into the quality of the COSs currently available and will support better COS use and development.

How to Measure Instructional Sensitivity? An Investigation of the Judgemental Approach

Keywords: Achievement, Accountability and school inspections, Effective instruction and teaching, Psychometrics
Presenting Author:Stephanie Musow, University of Teacher Education St.Gallen, Switzerland; **Co-Author:**Alexander Naumann, German Institute for International Educational Research (DIPF), Germany; **Co-Author:**Jan Hochweber, University of Teacher Education St. Gallen, Switzerland; **Co-Author:**Johannes Hartig, German Institute for International Educational Research (DIPF), Germany

Instructional sensitivity refers to the extent to which tests or test items are capable of capturing effects of instruction (Polikoff, 2010). Ensuring the instructional sensitivity of a test or a test item is important when students' test scores are used for teaching and school development, or teaching effectiveness research. If instruments are not instructionally sensitive, it isn't possible to draw valid inferences regarding teaching and school (Popham et al., 2014). In our contribution, we investigate whether a judgemental approach is appropriate for evaluating instructional sensitivity of test items. We conduct judgements of 32 math achievement items using (a) two different rating systems (global and differentiated) and (b) two different judgement groups (teachers and subject experts), and (c) compare the results to a more established test-data-based approach. The results show that (a) the differentiated and the global rating provide acceptable results, whereby the differentiated ratings shows slightly higher psychometric quality, (b) teachers seem to be the more reliable judges compared to subject experts, and (c) the results of the judgment and the test-data-based approach coincide in most cases with respect to the teacher but not to the subject expert. Still, teachers identified more items as instructional sensitive compared to the results of the test-based approach. Overall, our findings suggest that the judgemental approach appears promising. However, an adequate number of judges (about 10) is needed. Given that judgements do not need extensive test data, future research might investigate the usefulness of the judgemental approach for the construction of instructionally sensitive items.

Measuring teacher-student interaction in the classroom

Keywords: Classroom climate, Improvement, Longitudinal analyses/research, Quantitative methods
Presenting Author:Trude Havik, University of Stavanger, Norway; **Presenting Author:**Sigrun K. Ertesvåg, University of Stavanger, Norway

The increased interest in teacher's classroom interaction skills and its effects on students' social, motivation and academic outcomes bring up the question of evaluation approaches. Measuring teacher-student interaction in the classroom may also be of interest in the evaluation of strategies, interventions and initiatives to increase the quality of classroom interaction between teachers-students. The current study aims at illustrate and discuss the use of student reports and classroom observations to measure teacher-student interactions in the classroom. Strengths and limitations with both approaches are discussed. The study draws on data from 81 teachers and a classroom (1650 students) they were teaching at 14 schools with lower secondary grade students. The schools participated in a national initiative on developing lower secondary schools. Data from student surveys (two times) and classroom observations (four lessons) scored with the Classroom Assessment Scoring System Secondary (CLASS-S) were compared. Moreover, profile membership identified by Latent Profile Analysis in a previous study were used to identify subgroups of classrooms scoring high and low on all aspects of CLASS-S. For illustrations of similarities and differences in student reports and observation scores, the two subgroups were compared. Preliminary results indicate similar trends between observations and student reports. However, some interesting differences were found and will be elaborated on.

Session Wed 29, 14:00 - 15:30 3

29 August 2018 14:00 - 15:30

Auditorium
Symposium
Assessment and Evaluation

The Individual Level Revisited: Students' Idiosyncratic Perceptions of Teaching Quality

Keywords: At-risk students, Classroom climate, Effective instruction and teaching, Hierarchical linear modeling, Large scale assessment/large scale survey, Quantitative methods, Student background

Interest group: SIG 18 - Educational Effectiveness

Chairperson: Benjamin Caspar Fauth, University of Tübingen, Germany

Chairperson: Richard Goellner, University of Tuebingen, Germany

Organiser: Benjamin Caspar Fauth, University of Tübingen, Germany

Organiser: Richard Goellner, University of Tuebingen, Germany

Discussant: Tim Mainhard, Utrecht University, Netherlands

In recent years, research has made remarkable progress in the modeling of students' perceptions of teaching quality. However, most research used students' aggregated perceptions at the classroom level to assess teaching quality. Indeed, teaching quality is a classroom construct in nature. Accordingly, differences within classrooms, named idiosyncratic perceptions of students, have received much less attention. This is surprising, however, because students' idiosyncratic perceptions of teaching quality can theoretically be regarded as an important intermediary for understanding students' individual learning within specific learning contexts. The three contributions of this symposium ask for the conditions and consequences of individual students' perceptions from different perspectives.

The first paper examines whether students with different family backgrounds in terms of socio-economic status perceive the support of their teacher differently. Results show that high-SES students tend to evaluate their teacher more critically compared to students from lower SES families. The second paper examines the role of consensus as an indicator of classroom climate in student ratings. Results show that consensus within classes is very low for many items and that consensus is significantly associated with student characteristics and with classroom observation data. The third paper investigates if and how response tendencies shape student ratings of teachers' monitoring and learning support. The findings show that the association between response tendencies and students' perceptions of teaching was more pronounced for 5th graders than for 8th graders. Furthermore, mean-level differences in teaching quality ratings between the two age groups are substantially reduced after response tendencies are controlled.

Investigating consensus as an indicator of classroom climate in surveys of instructional practice

Presenting Author:Jonathan Schweig, RAND Corporation, United States; **Co-Author:**Jose Felipe Martinez, University of California, Los Angeles, United States; **Co-Author:**Meredith Langi, UCLA, United States; **Co-Author:**Jessica Schnittka, UCLA, United States

Student surveys are widely used to appraise classroom climate. Climate indicators derived from these surveys have demonstrated reliability and are predictive of student outcomes. This study examines whether indicators of consensus among student responses may also reflect features of the learning environment. Using math and science classroom data from the Measures of Effective Teaching (MET) project, we investigate whether perceptions of classroom climate are shared among students, whether there are systematic differences in student perceptions based on background characteristics, and whether indicators of consensus relate to student achievement and observation-based ratings of instructional practice. Results show perceptions are not strongly shared within classrooms, and that female students and low achieving students have less favorable impressions of classrooms than their peers. Consensus is significantly related to observation-based ratings of instruction. Implications of results for teacher feedback and evaluation are discussed.

Response Tendencies in Students' Individual Ratings of Teaching Quality

Presenting Author:Johanna Marder, University of Tuebingen, Germany; **Co-Author:**Richard Goellner, University of Tuebingen, Germany; **Co-Author:**Benjamin Caspar Fauth, University of Tübingen, Germany

Student ratings are a frequently used approach to gain information about the teachers' instructional quality in terms of classroom management, cognitive activation and instructional support. However, when answering questions on a rating scale, respondents may show preferences for certain response categories regardless of content, which could be a plausible explanation for why students from one and the same class disagree in their ratings. Analyzing data from a large sample of $N = 4172$ lower secondary school students, we examined whether student ratings of instructional quality are affected by response tendencies (i.e., net-acquiescence and extreme response style). To account for age related differences in students' response tendencies, we compared students from two age cohorts (5th and 8th grade students). Results showed that student ratings of teaching quality were clearly affected by students' response styles. Whereas 5th grade students' ratings showed effects of net-acquiescence and extreme response style, only net-acquiescence seemed to play a role in 8th grade students' ratings. Overall, effects for 5th grade students were more pronounced than for 8th grade students. Furthermore, mean-level differences for monitoring and learning support between 5th and 8th grade students were substantially reduced after response tendencies were controlled. The results indicate the importance to consider response tendencies in the analysis of student ratings of instructional quality, especially when asking students at younger ages.

Relationships between SES, Prior Achievement and Students' Perception of Teacher Support

Presenting Author:Cansu Atlay, LEAD Graduate School, Germany; **Co-Author:**Nicole Tieben, University of Tuebingen,

Germany; **Co-Author:**Steffen Hillmert, University of Tuebingen, Germany; **Co-Author:**Benjamin Caspar Fauth, University of Tübingen, Germany

This study examines the association between student socioeconomic background characteristics and their perception of teacher support (n=2243) and the interaction of prior achievement with this association. Building on the sociological work by Lareau (2003) and Calarco (2011), we suggest that there is a link between students' socioeconomic background and how they perceive support within the classroom. We expect that students from higher socioeconomic backgrounds would be more critical about the support they receive in the classroom and rate their teachers more negatively. Mixed effects models confirmed that, there is a negative association between parental educational attainment and wealth, and children's perception of teacher support, which was in line with the "concerted cultivation" theory by Lareau (2003). Having higher prior achievement was associated with a more positive perception of teacher support, but it didn't seem to interact with socioeconomic background.

Session Wed 29, 14:00 - 15:30 4

29 August 2018 14:00 - 15:30

Conference room 1

Symposium

Interventions smoothening students' transition from primary to secondary education

Keywords: Cognitive skills and development, Data-based decision making and data use, Developmental processes, Primary education, Program implementation and evaluation, Secondary education, Teacher professional development

Interest group: SIG 18 - Educational Effectiveness

Chairperson: Mieke Goos, UCLL, Belgium

Discussant: Leonidas Kyriakides, University of Cyprus, Cyprus

The transition from primary to secondary education is widely recognized as being challenging (Hanewald, 2013), for some students proceeding less smoothly, leading to study choice doubts, inclined achievement growth, a reduced school well-being, and/or grade retention (for reviews, see Hanewald, 2013; Hopwood, Hay, & Dymont, 2016; Korpershoek et al., 2016). The purpose of this symposium is to present findings from 3 different studies focusing on interventions easing the transition process among students: (1) a digital support system (and teacher training) to help teachers and students with students' text comprehension, self-regulation and motivation after the transition to secondary education (the Netherlands); (2) a joint social-emotional training and transitional program (Portugal); and (3) an across-educational level professional learning community focused on didactics (Belgium). Suggestions for educational research, practice, and policy will be discussed.

Transition to secondary education in the Netherlands: Effects of a digital support system

Presenting Author:Marie-Christine Opdenakker, University of Groningen, Netherlands; **Co-Author:**Marlies ter Beek, University of Groningen, Netherlands; **Co-Author:**Leonie Brummer, University of Groningen, Netherlands; **Co-Author:**Alma Spijkerboer, University of Groningen, Netherlands; **Co-Author:**Hidde Ozinga, University of Groningen, Netherlands

Promoting school success in the transition from elementary to middle school

Presenting Author:Vitor Alexandre Coelho, Académico de Torres Vedras, Portugal

Mind the gap! The impact of an across-educational level PLC focused on didactics

Presenting Author:Mieke Goos, UCLL, Belgium; **Co-Author:**Anne Decelle, UC Leuven-Limburg, Netherlands; **Co-Author:**Adelheid Spillier, UCLL, Belgium; **Co-Author:**Sonja Peeters, UCLL, Belgium

Session Wed 29, 16:00 - 17:00 1

29 August 2018 16:00 - 17:00

Auditorium

SIG 18 - 23 Keynote Session

SIG 18 - 23 Keynote Session "Designing Rigorous Evaluations of Educational Interventions"

Keywords: Experimental studies, Hierarchical linear modeling, Quantitative methods, Quasi-experimental research

Interest group:

Chairperson: Roel J. Bosker, Rijksuniversiteit Groningen, Dept of Education and GION, Netherlands

Designing Rigorous Evaluations of Educational Interventions

Keywords: Experimental studies, Hierarchical linear modeling, Quantitative methods, Quasi-experimental research

Presenting Author:Jessaca Spybrook, Western Michigan University, United States

Cluster randomized trials have become increasingly common designs to assess the impact of educational interventions. The process of planning and implementing a rigorous cluster randomized trial involves multiple steps. This talk will focus on three of those steps:

1) strategies for selecting the appropriate design,

2) given the specific design, steps for conducting power calculations to answer the *What Works and For Whom or Under What Conditions* questions, and

3) steps for increasing the credibility and transparency of the impact study through pre-registration of the design and analysis plan.

I will discuss the design considerations and power calculations using real examples from impact studies of educational interventions. I will demonstrate the features of pre-registration using the Registry of Efficacy and Effectiveness Studies in Education (REES), a registry specifically targeted at impact studies in education. REES was developed with support from the Institute of Education Sciences, the research branch of the U.S. Department of Education, and is set to launch in 2018.

Session Thu 30, 09:30 - 11:00 1

30 August 2018 09:30 - 11:00

Leekstermeerzaal

Single Paper

Teaching and Teacher Education

Single Paper Improvement teacher level

Keywords: Class and school composition, Effective instruction and teaching, Experimental studies, Learning communities, Longitudinal analyses/research, Program implementation and evaluation, Secondary education, Self-efficacy, Teacher professional development

Interest group: SIG 18 - Educational Effectiveness, SIG 23 - Educational Evaluation, Accountability and School Improvement

Chairperson: Jana Strakova, Faculty of Education, Charels University in Prague, Czech Republic

Schools' demographic characteristics and support activities provided to beginning teachers

Keywords: Class and school composition, Program implementation and evaluation, Teacher professional development, Secondary education

Presenting Author: Marieke van der Pers, University of Groningen, Netherlands; **Co-Author:** Michelle Helms-Lorenz, University of Groningen, Netherlands

Teacher induction is a widespread instrument used to attract, develop and commit beginning teachers (BTs) to teaching. BTs differ in their ability to teach and in the rate they develop their teaching skills during the first years on the job (Maulana, Helms-Lorenz & Van de Grift, 2015). Schools differ in the amount and intensity in which they (are able to) offer support to beginning teachers. This study explores the extent to which schools' sociodemographic characteristics relate to the provided support as reported by BTs. The aim is to find room for improvement of induction arrangements. The data were drawn from 3 cohorts of a national induction programme in secondary education in the Netherlands in which qualified BTs are supported in their first years of their professional practice. A support monitor measures BTs' (N=1,716) perception of the support provided by the school (N=158). The monitor consists of questions concerning four induction arrangement elements administered at the end of the school years. For the four program elements differences between schools concerning the proportion of teachers being 55 years and older are found. Preliminary results from multinomial regression analyses also indicate that with an increasing proportion of teachers being 55 years and older, schools are less likely to offer an more than average number of support activities. Future research should aim to extent the knowledge of how the presence of 'older' experienced teachers affects the implementation process of induction arrangements in schools.

Feedback, written or video? The possible impact on secondary education

Keywords: Effective instruction and teaching, Self-efficacy, Secondary education, Experimental studies

Presenting Author: Esther Canrinus, University of Agder, Norway; **Co-Author:** Anne-Line Bjercknes, University College of Southeast Norway, Norway

Research in higher education finds that video-feedback strengthens the bond between the teacher and the student. We explore, in secondary education, the impact of video-feedback on pupils' perception of teaching, and the impact of video-feedback on pupils' self-efficacy for self-regulated learning. In our cross-over design study, four secondary school classes participated. In the first semester (T1), half of one class received video feedback and the other half received written feedback. This switched during the second semester (T2). Three different classes participated as control groups. Mean age was 16 at T1. At T1 and T2, pupils completed a survey on self-efficacy (N_{T1}=108; N_{T2}=96). At T2, additional items on their teachers' teaching and their own engagement were included. We used the Kruskal-Wallis test to explore differences. We found no significant differences between the feedback-groups at T1. At T2, the control group expressed significantly more self-efficacy for self-regulated learning compared to the group receiving written instruction (p.05). No differences in grades were observed. An explanation to why we find no differences between groups might be the short time this feedback procedure has been implemented. Yet, our study sheds light on the impact of video-feedback on various aspects of teaching. At the conference we will present data (collected spring 2018) on teachers' experiences. This will give insight into the pros and cons of using video-feedback.

Comparing the influence of teacher development on student outcomes in China and England

Keywords: Effective instruction and teaching, Learning communities, Teacher professional development, Longitudinal

analyses/research

Presenting Author:Sally Mary Thomas, University of Bristol, United Kingdom; **Co-Author:**Lei Zhang, University of Bristol, United Kingdom

This paper compares selected new findings from two separate studies conducted in China and England which examine the influence of Professional Learning communities and teacher development on student outcomes. The first study funded by the UK Educational Endowment Foundation (EEF) employed RCT methodology to evaluate differences in student outcomes between English schools that participated over two years in a Research Learning Community intervention and control schools that did not. The second study funded by UK DFID/ESRC collected largescale datasets to examine the observed association between Professional Learning Communities (PLCs), teacher development variables and student outcomes in China. The paper seeks to provide new insights about the impact of teacher development and other factors on students' attainment and progress at school in two country contexts as well as review the potential role of research design and context specificity in shaping the findings. In both studies innovative quantitative methodology (multilevel modelling) is used to analyse longitudinal datasets comprising students' examination, prior attainment and background information as well as teacher/school process and context data from 100+ schools. The findings from English schools did not identify a statistically significant impact of the Research Learning Communities intervention on student "value added" outcomes at KS2, whereas in China some statistically significant links were identified between teacher development and senior secondary school students' "value added" outcomes. The findings are discussed in relation to teacher development and school improvement in different contexts, evaluation methodology and the implications for school evaluation and accountability.

Session Thu 30, 09:30 - 11:00 2

30 August 2018 09:30 - 11:00

Conference room 2

Single Paper

Teaching and Teacher Education

Single Paper Ratings of teaching quality 2

Keywords: Achievement, Effectiveness and efficiency, Improvement, Parental involvement in learning, Program implementation and evaluation, Psychometrics, Quantitative methods, Secondary education, Self-efficacy, Teacher professional development

Interest group: SIG 18 - Educational Effectiveness

Chairperson: Rikkert van der Lans, University of Groningen, Netherlands

Factors influencing student perceptions of teaching quality

Keywords: Effectiveness and efficiency, Teacher professional development, Secondary education, Psychometrics

Presenting Author:Hannah Bijlsma, Fac. of Behavioural Sciences, University of Twente, Netherlands; **Co-Author:**Adrie Visscher, Univ. of Twente, Unknown

Teaching quality can be measured in several ways. Lesson observations are quite common, however, measuring the perceptions of the target group, the students, about teaching quality is not done very frequently yet. Modern technology enables the efficient collection and processing of student perceptions of teaching quality as well as feeding back the results to teachers in order to support the improvement of teaching quality. For that purpose, the digital Impact! tool was developed. Although student perceptions become increasingly prominent in the evaluation of teachers, the validity of students' teaching quality judgements may be questioned as ratings may be biased. For example, because of the interest in and performance of students in the subject taught by the teacher they rate, and the general popularity of the teacher who is rated. Research into the validity of student perceptions of teaching quality is still very limited and insight into how much the factors mentioned influence student perceptions of teaching quality is lacking. In order to fill this gap in our knowledge this study investigated the validity of student perceptions of teaching quality by investigating which student, class, teacher and time characteristics influence those perceptions. The results clarify which factors should be taken into account when using student feedback.

Does perceived parents' and teacher's support affect students' self-efficacy?

Keywords: Achievement, Parental involvement in learning, Self-efficacy, Quantitative methods

Presenting Author:Josip Sabic, Institute for Social Research, Croatia; **Co-Author:**Jelena Matić, Institute for Social Research - Zagreb, Croatia, Croatia; **Co-Author:**Iva Košutić, Institute for Social Research - Zagreb, Croatia, Croatia; **Co-Author:**Zrinka Ristic Dedic, Institute for Social Research, Croatia

Effective schools should not only enable students to acquire knowledge and skills. They should help students to develop a positive self-regard, as well as an optimistic attitude when faced with the multiplicity of academic tasks. Among the important actors that stimulate or hinder the development of the academic self-efficacy in students, parents and teachers play a prominent role. High levels of parental interest and engagement in their children's schooling, together with the caring approach of the teacher might be especially relevant for the development of the younger students' feelings of self-efficacy. This paper examines how predictive are the perception of parents' and teacher's support for the students' self-efficacy

reports, over and above their estimated academic achievement. The quantitative study encompassing these measures was conducted on the sample of 1062 fourth grade students in the city of Zagreb, Croatia. Data was analysed using hierarchical regression analysis. Estimated students' achievement explained around 14 % of the variance of the students' self-efficacy. The perception of parents' and teacher's support explained unexpectedly low percent of the students' self-efficacy above the information about their school success (6 %). Overall, the indicators explained around 20 % of the variance of the students' self-efficacy, with estimated achievement being most predictive. The perception of parents' support was somewhat weaker predictor of the students' self-efficacy than the perception of teacher's support. The results are discussed in light of the need of parental and teacher support to become more empowering and long-term oriented, versus more operative, task-oriented and short-term.

Teachers' usage of student feedback. A typology.

Keywords: Improvement, Program implementation and evaluation, Teacher professional development, Secondary education

Presenting Author:Sebastian Röhl, University of Education Freiburg, Germany; **Co-Author:**Wolfram Rollett, University of Education Freiburg, Germany

Several scholars describe feedback from students to teachers as a useful instrument for the professional development of teachers and the improvement of instruction. At the same time, empirical studies show a heterogeneous picture of the use of student feedback by teachers. This study uses cluster analysis to identify different types of teachers in their dealing with student feedback. The sample consists of 98 teachers of an association of German schools. Analysis included the utilization of the received student feedback by teachers, their attitude towards student feedback and the perception of the feedback received from their students. Five stable types could be identified, including 1) consistent positive users, 2) critical users, 3) users with a negative perception as well as 4) non-users with positive attitudes and 5) non-users with negative attitudes. Further analysis of the types focusing on occupational stress and coping behaviour emphasize a high relevance of teachers' coping capacities for their dealing with the student feedback. Indications for the implementation and support of student feedback procedures at schools will be discussed.

Session Thu 30, 09:30 - 11:00 3

30 August 2018 09:30 - 11:00

Conference room 1

Single Paper

Teaching and Teacher Education

Single Paper Interactions 2

Keywords: Classroom climate, Effective instruction and teaching, Effectiveness and efficiency, Improvement, Quantitative methods, Secondary data analysis, Teaching approaches, Video analysis

Interest group: SIG 18 - Educational Effectiveness, SIG 23 - Educational Evaluation, Accountability and School Improvement

Chairperson: Nicolas Hübner, University of Tübingen, Germany

Identifying inspiring versus effective teaching: How do they link and differ?

Keywords: Effectiveness and efficiency, Teaching approaches, Quantitative methods, Secondary data analysis

Presenting Author:James Ko, The Education University of Hong Kong, Hong Kong; **Co-Author:**Pamela Sammons, University of Oxford, United Kingdom; **Co-Author:**Ridwan Maulana, GION - University of Groningen, Netherlands; **Co-Author:**Wanlu Li, The Education University of Hong Kong, Hong Kong; **Co-Author:**Leonidas Kyriakides, University of Cyprus, Cyprus

A vast amount of literature on teacher effectiveness has accumulated, but that on inspiring teaching is very scarce. Many characteristics of inspiring teachers proposed or identified are often similar to those found in effective teachers. The theoretical boundaries between inspiring and effective teaching are often overlapped. Accordingly, this study attempted to develop a new instrument to measure inspiring teaching, based on a theoretical framework proposed by Sammons et al. (2014) on inspiring teachers. Inspiring teaching is hypothesised to share similar behavioural characteristics with effective teaching to some extent (generic), while certain characteristics are assumed to refer solely to inspiring teaching (specific). A quantitative instrument with 68 items covering behaviours in 12 teaching and learning aspects was developed using the Delphi method. Consistent with the hypotheses, the results indicated that eleven teaching-related factors clustered in two groups, suggesting that factors including flexibility, reflectiveness, and innovative teaching were more likely to be related to inspiring teaching while the others were associated with effective teaching. Inspiring teaching was also found in the early stage of teaching and effective teachers were not always inspiring. Thus, some professional development programs can be developed to support teachers with to adopt more innovative and teaching flexibility. This study has contributed to our understanding of the similarities and difference in inspiring and effective teaching and offered an instrument for researchers to study inspiring teaching in other contexts.

Classroom interaction –teacher perceptions and classroom observations

Keywords: Classroom climate, Effective instruction and teaching, Improvement, Video analysis

Presenting Author:Sigrun K. Ertesvåg, University of Stavanger, Norway; **Presenting Author:**Grete Sørensen Vaaland, University of Stavanger, Norway

Teachers' interactions with their students have been described in terms of teachers' emotional support, classroom organization and instructional support. Teacher-student interactions in the classroom are among the most important factors in providing a supportive environment for the academic and social emotional learning of students. Numerous professional development efforts to improve teacher-student interaction are initiated as part of teacher-, school-, district- or national level initiatives to improve teaching practices. Teachers' improvement in practice through these initiatives are rarely evaluated. When evaluated, mostly through self- or student reports. The rapid technological development has made video observation more available, at the same time high quality observation scoring systems have been developed. The current study investigate similarities and differences in level and variation of scores of teachers self-reported emotional support, classroom organization and instructional support and corresponding classroom observations scores. The sample contained 81 teachers and a classroom they were teaching (1650 students). The teachers reported on a web-based questionnaire in the beginning and the end of one academic year. In addition, classrooms were observed four lessons throughout the year. The observations were scored on 11 dimensions using Classroom Assessment Scoring System-Secondary (CLASS-S). Profile membership, identified in a previous study, of teacher-student interaction profiles of 11 dimensions were used as a starting point. Although level and variation in scores differed, preliminary analysis indicates that there are similar trends in results from teachers self-reports and observations between teachers with different profile membership. Methodological implications are discussed with respect to evaluation of teaching practices.

Session Thu 30, 09:30 - 11:00 4

30 August 2018 09:30 - 11:00

Auditorium

Symposium

Teaching and Teacher Education

Better Together? Merits and Limitations of Researching Instructional Quality More Synergistically

Keywords: Effective instruction and teaching, Primary education, Secondary education, Teaching approaches

Interest group: SIG 18 - Educational Effectiveness

Chairperson: Charalambos Charalambous, University of Cyprus, Cyprus

Chairperson: Anna-Katharina Praetorius, German Institute for International Educational Research (DIPF), Germany

Discussant: Daniel Muijs, University of Southampton, United Kingdom

Over the past decades, researchers working on instructional quality have used a multitude of approaches for capturing instructional quality. Looking back, it is obvious that these studies enabled us to learn a lot about instruction (for an overview, see e.g., Seidel & Shavelson, 2007). At the same time, prior attempts appear to have been diverse so that moving the field forward by building upon each other's work seems possible only to a limited degree. Looking forward, we argue that we need to join efforts and intensify our attempts to learn from each other by starting to see different frameworks synergistically. Toward this end, in this symposium a synthesis of instructional aspects that resulted from reviewing 12 classroom observation frameworks (content-generic, content-specific, and hybrids) is first presented. By inviting scholars representing content-generic frameworks and content-specific frameworks to react to this synthesis, we then intend to carve out the merits and limitations of the synthesized work compared to individual frameworks as well as point to needs of further development of such collaborative efforts. The symposium's discussion critically examines the theoretical, methodological, and practical promises and the shortcomings of such efforts and provides ideas for fruitful next steps in the attempt to work more collaboratively in instructional research. From this viewpoint, the symposium resonates well with the theme of this year's conference, in that it seeks to provide an arena for discussing how by joining efforts instructional quality can better be conceptualized, understood, measured, and eventually, improved.

Commonalities and Differences across Classroom Observation Frameworks for Studying Teaching Quality

Presenting Author:Anna-Katharina Praetorius, German Institute for International Educational Research (DIPF), Germany;

Co-Author:Charalambos Charalambous, University of Cyprus, Cyprus

Researchers from different fields have developed different observation-based frameworks to capture instructional quality with a focus on content-generic versus content-specific dimensions or combinations thereof (i.e., hybrid frameworks). These frameworks differ in what they capture, but also in the way they structure and name certain instructional aspects, making it hard to generate accumulated knowledge across frameworks. The need to explore synergies and complementarities among existing works on instruction and its quality therefore becomes imperative. The current paper aims at outlining the work of synthesizing different aspects of instructional quality emanating from 12 observation-based frameworks using an inductive qualitative approach. The elements that stemmed from this work were grouped into seven dimensions. An analysis of these dimensions reveals that (a) clear patterns among the three categories of frameworks (i.e., content-generic, content-specific, and hybrid) exist, (b) considerable differences among frameworks even within the same category can be observed, and (c) some aspects appear equally across content-generic and content-specific frameworks. The key finding, however, is that the frameworks can function complementarily since none of them appears to capture all aspects of instructional quality identified across frameworks. Still, large overlaps could be found between frameworks, even

for those from different categories. We thus wonder whether indeed so many frameworks are needed or whether it would be wiser to bundle our efforts. We discuss the theoretical and practical implications of this work.

Towards a Generic Theory of Teaching Quality: Benefits of Integrative Frameworks and Meta-Analyses

Presenting Author:Eckhard Klieme, German Institute for International Educational Research (DIPF), Germany; **Co-Author:**Leonidas Kyriakides, University of Cyprus, Cyprus

The ultimate challenge for educational science is to develop a theory of effective teaching that is applicable to a broad variety of content areas and classroom settings, even across pedagogical traditions and cultures. This task requires theories and methods from various strands of research to be combined through any of the following ways: 1) Setting up an eclectic taxonomy of "effective teaching", 2) accumulating empirical evidence through meta-analyses, 3) theoretical accumulation through careful and precise inspection of constructs and measures from existing frameworks (Praetorius and Charalambous, this symposium), 4) reducing a broader set of measures to a more parsimonious model of basic dimensions through empirical analysis (as Klieme et al did in TIMSS-Video/Germany), 5) drawing a small set of factors from theories of effective teaching within a broad theory of schooling as in Creemers' and Kyriakides' "Dynamic Model of Educational Effectiveness". The present paper is based on approaches # 4 and 5 and responds to the questions guiding this symposium: what are the merits, limitations, and prospects of a more comprehensive approach like #3? We claim that generic frameworks are important drivers of theory development, as they allow to study teaching across multiple contexts. However, there are blind spots which may be enlightened by combining multiple frameworks.

Merits and Constraints of Mathematics-Specific Frameworks and General Indicators of Teaching Quality

Presenting Author:Erica Litke, University of Delaware, United States

In this paper, researchers from the development teams of three mathematics-specific classroom observation protocols consider the charge issued by the symposium organizers: how can individual projects work more synergistically to better conceptualize, understand, measure, and improve instructional quality? We address the affordances and constraints of analyzing mathematics lessons using a combined list of indicators of teaching quality as compared to using mathematics-specific instruments. In addition, we look across mathematics-specific instruments to highlight the focus and purpose of classroom observation from a research and instructional improvement perspective.

Session Thu 30, 11:30 - 12:30 1

30 August 2018 11:30 - 12:30

Auditorium

SIG 18 - 23 Keynote Session

SIG 18 - 23 Keynote Session "Success for All: Schoolwide Improvement for Disadvantaged Students"

Keywords: Educational reform, Effective instruction and teaching, Parental involvement in learning, Teacher professional development

Interest group:

Chairperson: Roel J. Bosker, Rijksuniversiteit Groningen, Dept of Education and GION, Netherlands

Success for All: Schoolwide Improvement for Disadvantaged Students

Keywords: Educational reform, Effective instruction and teaching, Parental involvement in learning, Teacher professional development

Presenting Author:Robert Slavin, Johns Hopkins University, United States

Success for All is a whole-school approach designed to use all school staff and resources in an integrated way to ensure that all students in elementary and middle schools succeed, especially in reading. Begun in 1987 in Baltimore, it is now used widely in the US and UK, and on a smaller scale in the Netherlands. The main components of the approach include cooperative learning, extensive professional development, frequent assessment, tutoring for struggling students, parent involvement, and a whole-school leadership structure. Numerous evaluations, mostly by third parties, have found positive effects of Success for All on reading. This presentation will explain the development and current elements of Success for All, the research, the dissemination, the way this program leads to sustainable school improvement, and the place of this strategy in the evidence based reform movement in the US and UK.

Session Thu 30, 13:30 - 15:00 1

30 August 2018 13:30 - 15:00

Leekstermeerzaal

Single Paper

Assessment and Evaluation, Educational Policy and Systems

Single Paper Assessment

Keywords: Achievement, Attitudes and beliefs, Cognitive skills and development, Educational attainment, Educational policy, Effective instruction and teaching, Effectiveness and efficiency, Experimental studies, Longitudinal analyses/research, Psychometrics, Quantitative methods

Interest group: SIG 01 - Assessment and Evaluation, SIG 18 - Educational Effectiveness

Chairperson: Hannah Bijlsma, Fac. of Behavioural Sciences, University of Twente, Netherlands

Validity of Test Score Interpretations: The Role of Instructional Sensitivity

Keywords: Achievement, Effective instruction and teaching, Longitudinal analyses/research, Psychometrics

Presenting Author:Alexander Naumann, German Institute for International Educational Research (DIPF), Germany; **Co-Author:**Stephanie Musow, University of Teacher Education St.Gallen, Switzerland; **Co-Author:**Svenja Rieser, Bergische Universität Wuppertal, Germany; **Co-Author:**Jan Hochweber, University of Teacher Education St. Gallen, Switzerland; **Co-Author:**Johannes Hartig, German Institute for International Educational Research (DIPF), Germany

Assessments of students' achievement or competencies are widely used in educational research and policy to draw inferences on the effectiveness of teaching. Yet, valid inferences on teaching effectiveness drawn from students' test scores require that instruments are capable of capturing effects of classroom instruction, that is, instructionally sensitive. For example, if the instructional sensitivity of a test is unknown prior to a study, interpretations whether a lack of effects is due to ineffectiveness of teaching or insensitivity of the test remain inextricably confounded. Instructional sensitivity is defined as the psychometric property of a test or a single item to capture effects of the content and the quality of classroom instruction. Although there are various statistical approaches to measure the instructional sensitivity of items, the measures' relationship to actual classroom instruction is rather unclear. In addition, there is little knowledge on how item sensitivity relates to the test's sensitivity. The present study aims at closing these gaps by investigating test and item sensitivity to teaching quality. More specifically, we evaluate the instructional sensitivity of test items and test scores to the degree of cognitive activation in class, and the consequences for test score interpretation when a test is assembled from items varying in their degree of sensitivity to cognitive activation. Our study (a) provides validity evidence that item sensitivity measures may be related to actual classroom instruction and (b) points out that inferences on teaching drawn from test scores may vary due to test composition.

Effects of Interim Assessments on the Achievement Gap

Keywords: Educational policy, Effectiveness and efficiency, Experimental studies, Quantitative methods

Presenting Author:Spyros Konstantopoulos, Michigan State University, United States; **Co-Author:**Wei Li, University of Alabama, United States; **Co-Author:**Shazia Miller, NORC, United States; **Co-Author:**Arie van der Ploeg, AIR, United States

We analyze data from a field experiment conducted in K-8 public schools in Indiana in the U.S. in 2010-2011 to examine the effects of two interim assessment programs (*mCLASS* in grades K-2 and *Acuity* in grades 3-8) on mathematics and reading achievement. We employ quantile regression to examine the consistency of the treatment effects in the middle, the lower, and the upper tails of the achievement distribution. Results indicate that treatment effects are not significant across the achievement distribution in grades 3-8 indicating that *Acuity* did not have an impact on student achievement. In grades K-2 however, there is a negative and significant effect of *mCLASS* on student achievement in the middle and the tails of the achievement distribution. The effects of interim assessments are consistent across the achievement distribution, which suggests that interim assessments do not have an impact on the achievement gap between lower- and higher-achievers.

What counts for achievement – Time on task in low and high stakes assessment

Keywords: Achievement, Attitudes and beliefs, Cognitive skills and development, Educational attainment

Presenting Author:Sirkku Kupiainen, University of Helsinki, Finland; **Co-Author:**Risto Hotulainen, University of Helsinki, Finland

Computer-based assessment (CBA) has allowed the adoption of response time (RT) as an indicator of student effort in addition to or instead of a questionnaire, and a key prerequisite for the reliability and validity of assessment results – in class as well as in international comparative studies. RT relates to Carroll's (1963) notion of time-on-task (TOT), crucial for all learning. The present study incorporates two strands of educational research and theory: The relative impact of cognitive competence and motivational factors in explaining achievement, and the mediating role of TOT on students' attainment in high-stakes vs. low-stakes assessment. Additionally, the study allows the replication and validation of an earlier study, carried out five years earlier in the same schools and with the same instruments, indicating this mediation in a low-stakes assessment. The data is drawn from a national sample of 7780 Finnish ninth graders in an assessment of curricular and non-curricular competences and motivation. It was hypothesised – and shown by the results of the study, using structural equation modelling – that TOT mediates the role of earlier achievement (GPA) and of students' motivational attitudes in predicting new achievement. The role of TOT was stronger in low-stakes assessment and stronger for boys than for girls. It was also shown that the relative roles of GPA, motivation, and TOT in explaining students' attainment in the low stakes non-curricular test have stayed approximately the same as they were five years earlier. The results have implications for schools/ teachers and for interpreting the results of international assessments.

Session Thu 30, 13:30 - 15:00 2

30 August 2018 13:30 - 15:00

Conference room 2

Single Paper

Learning and Social Interaction, Motivational, Social and Affective Processes

Single Paper Affective measures

Keywords: Attitudes and beliefs, Early childhood education, Emotion and affect, Longitudinal analyses/research, Motivation, Primary education, Program implementation and evaluation, Secondary education, Self-efficacy, Self-regulation, Student background

Interest group: SIG 18 - Educational Effectiveness

Chairperson: James Ko, The Education University of Hong Kong, Hong Kong

Healthy Minds, Healthy Schools: Effectiveness of a School-Based Emotion Regulation Program

Keywords: Emotion and affect, Program implementation and evaluation, Self-regulation, Early childhood education

Presenting Author: Tina Montreuil, McGill University, Canada; **Co-Author:** Micah Tilley, McGill University, Canada

The research literature clearly indicates that approximately 20-25% of children and adolescents have diagnosable mental health problems and 80% do not receive adequate treatment. Research has also shown that a child's behavioural and emotional regulation is a predictor of a child's future social and academic competence. Factored into this situation are the long waitlists in the community, limited mental health services in schools. Schools are where children and youth are likely to receive mental health services through the promotion of positive mental health and overall wellbeing and it was moreover determined that schools are the most effective place to provide mental health services. This is due to the fact that schools are familiar to families, and thus mental health stigma that prevents many to access these services; hence, supporting the need for youth mental health literacy in schools. Following a policy-oriented approach, Healthy Minds, Healthy Schools, a universal curriculum-based program, was implemented in schools through guided instruction, sections on personal reflection, and group activity exercises. The 16 lesson program included themes such as emotional awareness and regulation; mindfulness and metacognition; the linking thoughts and feelings; problem solving, etc. Regression analyses, independent t-tests and ANCOVA's, using a randomized control trial design, suggested that the program resulted in improved mental health outcome and wellbeing and improved student functioning and achievement. Findings validate the effectiveness of a universal school-based mental health program. Furthermore, the findings support the need for universal curriculum-based programs and policy-oriented approaches in promoting school-based mental health and wellbeing.

Ameliorating teacher expectations: Can student beliefs overcome low teacher expectations?

Keywords: Attitudes and beliefs, Self-efficacy, Student background, Primary education

Presenting Author: Christine Rubie-Davies, University of Auckland, New Zealand; **Co-Author:** Penelope Watson, The University of Auckland, New Zealand

Teacher expectations can have self-fulfilling prophecy effects on student achievement. However, this relationship is dependent on students assimilating, and acting according to teachers' expectations. Some students achieve well despite teachers' low expectations. This study explored student beliefs about intelligence and self-efficacy as possibly mediating the effects of low teacher expectations. Controlling for achievement, students were grouped according to over- or under-estimation of their achievement by teachers. Low expectation students scored higher in mathematics if they attended a high socioeconomic school or they had high self-efficacy. Teachers had lower expectations for boys and Māori than was warranted and over-estimated high SES and Asian students. The results emphasize the need for teachers to realize the importance of accurate student assessment and promotion of positive student self-beliefs.

The psychosocial development of Flemish students in the first two years of secondary education

Keywords: Attitudes and beliefs, Motivation, Secondary education, Longitudinal analyses/research

Presenting Author: Naomi Van den Branden, KU Leuven - University of Leuven, Belgium

This study investigates the development of academic self-concept, motivation, well-being and school engagement from Grade 7 to Grade 8 by repeated assessment of 5,614 Flemish students in 46 secondary schools. Multivariate multilevel models show a decline in all non-cognitive outcomes, except general academic self-concept. Furthermore, student background characteristics such as gender, socio-economic status, language at home, and prior achievement scores are related to the (evolution of) all non-cognitive outcomes. Finally, analyses show that schools do have a small but statistically significant effect on the (evolution of) these non-cognitive outcomes, although these differences between schools can partially be explained by student background characteristics.

Session Thu 30, 13:30 - 15:00 3

30 August 2018 13:30 - 15:00

Conference room 1

Symposium

Assessment and Evaluation

Data comparability as prerequisite for effective evaluations of learning and teaching

Keywords: Achievement, Attitudes and beliefs, Cross-national/cultural comparisons/International comparisons, Effectiveness and efficiency, Large scale assessment/large scale survey, Large scale assessment/large scale survey, Longitudinal analyses/research, Motivation, Quantitative methods, Teaching approaches

Interest group: SIG 18 - Educational Effectiveness

Chairperson: Jessica Fischer, German Institute for International Educational Research (DIPF), Germany

Discussant: Eckhard Klieme, German Institute for International Educational Research (DIPF), Germany

A major interest of educational effectiveness research is explaining why and to what extent student outcomes differ across groups (e.g. educational tracks or educational systems) as well as comparing their relationship with each other and with potential predictors (e.g. teaching quality) across groups. However, group-membership and contextual factors may shape the understanding of the constructs studied and thus influence the way survey items are interpreted, considered and ultimately answered. In case of systematic variation - especially likely in large scale studies involving heterogeneous populations (e.g. international assessments) - data comparability can seriously be jeopardized. Thus, before valid comparative inferences can be drawn, it is vital to demonstrate comparability of target measures to ensure that variation between groups lies in the targeted construct rather than being caused by non-equivalence of measures or response bias. Similarly, comparability across time points should be checked in trend studies and longitudinal studies. Despite the critical relevance, data comparability is hardly formally tested in large-scale educational assessments and longitudinal studies. This symposium gathers four examples of methodological issues as well as empirical applications relating to measurement invariance in the context of teaching and learning, with a focus on reading-related attitudes and perceived teaching quality. We showcase applications of various invariance testing methods (e.g., two-level multigroup confirmatory factor analysis, alignment method) and check whether item design (e.g., anchoring vignettes, negatively oriented items) may help improve data comparability. Measurement invariance is estimated over time as well as across groups (e.g. countries).

Attitudes Towards Reading: Analysis of measurement invariance over time and across levels

Presenting Author:Luisa Grützmaker, German Institute for International Educational Research (DIPF), Germany; **Co-Author:**Svenja Vieluf, German Institute for International Educational Research (DIPF), Germany; **Co-Author:**Carmen Köhler, German Institute for International Educational Research (DIPF), Germany; **Co-Author:**Johannes Hartig, German Institute for International Educational Research (DIPF), Germany

Many studies on school effectiveness use achievement as a relevant effectiveness criterion. However, this practice has been criticized and additional criteria were proposed, among them reading attitudes. In order to use reading attitudes as an effectiveness criterion in a longitudinal study, it is necessary to investigate whether measures of reading attitudes are comparable across time and different school types. Since attitudes are measured at the student level via questionnaires but are also used as indicators to inform about attitudes at the school level, it is further necessary to assure that the reading attitudes scale is an adequate scale at the individual and at the school level. Investigating measurement invariance of reading attitudes (1) across time, (2) levels and (3) school types was the purpose of this study. The study used longitudinal data from the KESS-study. The sample comprised 12.450 students in 172 secondary schools. Measurement invariance over time was tested by comparing 3 increasingly restrictive models. Subsequently, the most restrictive model that could be accepted in these analyses provided the basis for analyzing measurement invariance across time points and levels. Finally, the most restrictive model that could be accepted in these analyses was used to test measurement invariance across school types. The results indicate that scalar invariance over time, across levels and school types could be established. Consequently, the change of attitudes over time and across levels was captured adequately in different school types, and latent means across time, levels and between school types were comparable.

Self-concept in reading: subcomponents, cross-cultural invariance and relationships with achievement

Presenting Author:Dominique Lafontaine, Université de Liège, Belgium; **Co-Author:**Virginie Dupont, University of Liège, Belgium; **Co-Author:**Patricia Schillings, University of Liège, Belgium; **Co-Author:**Doriane Jaegers, University of Liège, Belgium

Using secondary data from 48 countries, this study investigated the factorial structure and tested the cross-cultural invariance of the PIRLS 2011 reading self-concept scale and the relationships between self-concept and reading achievement. The study showed that a two-factorial structure of the self-concept scale in reading is the most adequate model. As expected, both configural invariance and metric invariance were achieved, while scalar invariance was not. The two self-concept dimensions Perception of competence and Perception of difficulty showed robust within-country correlations with reading achievement, especially the second one. At the country level, the correlation of the Perception of competence with reading achievement was negative, illustrating the attitudes-achievement paradox: countries in which students on average reported a more positive self-concept performed lower on average. On the contrary, for the Perception of difficulty, the sign of the correlation remained the same at the within-country and at the country level. The attitudes-achievement paradox was not observed. These results support the value of including negatively oriented items (Perception of difficulty) in reading self-concept scales, especially in comparative large-scale assessments in which response-styles are likely to jeopardize the measurement of the targeted constructs.

Comparability of items measuring instructional quality across and within three language groups

Presenting Author:Jessica Fischer, German Institute for International Educational Research (DIPF), Germany; **Co-Author:**Anna-Katharina Praetorius, University of Zurich, Switzerland; **Co-Author:**Eckhard Klieme, German Institute for International Educational Research (DIPF), Germany

Using secondary data from PISA 2012, this study investigated the comparability of items measuring instructional quality in mathematics within and across three language groups as well as the influence of linguistic similarity on cross-cultural

equivalence. The main findings based on results of the alignment method showed that cross-cultural comparability, on the one hand, depends on the targeted construct and as expected, is positively influenced by linguistic similarity. The amount of non-invariant item intercepts was significantly lower for Constructivist Support, Classroom Management and Cognitive Activation when comparing three German- respectively three Spanish-speaking countries in contrast to comparing nine countries, not all sharing a common language. The degree of comparability of items measuring instructional quality, however, was much lower for the Chinese-speaking group in contrast to the German-and Spanish-speaking language-group. The Chinese-speaking group is the most heterogeneous with regard to language (e.g. Mandarin vs. Shanghai dialect) and cultural background (different colonial history), which might explain the lower degree of data comparability within this group. With respect to comparability according to construct, Cognitive Activation seems to be the most problematic with a higher amount of non-invariant intercepts compared to Constructivist Support and Classroom Management both within and across the language-groups. Thus, this study provides evidence, that comparing many heterogeneous groups resulting in non-comparability of latent means might ignore the fact that comparisons are possible for subgroups with identical or similar linguistic and cultural context.

Effects of Anchoring Vignettes on Comparability and Predictive Validity of Student Self-Reports

Presenting Author:Jia He, TilbuDIPF, China; **Co-Author:**Janine Buchholz, DIPF Frankfurt, Germany; **Co-Author:**Eckhard Klieme, German Institute for International Educational Research (DIPF), Germany

Anchoring vignettes are item batteries especially designed for correcting responses that might be affected by incomparability. This paper investigates the effects of anchoring vignettes on the validity of student self-report data in 64 cultures. Using secondary data analysis from the 2012 Program for International Student Assessment (PISA), we checked the validity of ratings on vignette questions, and investigated how rescaled item responses of two student scales: Teacher Support and Classroom Management enhanced comparability and predictive validity. The main findings include that (1) responses to vignette questions represent valid individual and cultural differences; in particular, violations in these responses (i.e., misorderings) are related to low socioeconomic status and low cognitive sophistication; (2) the rescaled responses tend to show higher levels of comparability; and (3) the associations of rescaled Teacher Support and Classroom Management with math achievement, student-oriented and teacher-directed instruction are slightly different from raw scores of the two target constructs, and the associations with rescaled scores seem to be more in line with the literature. Namely, the associations among all self-report Likert scales are weaker with rescaled scores, presumably reducing common method variance; and both rescaled scale scores are more positively related to math achievement. The country ranking also changes substantially; in particular Asian cultures top the ranking on Teacher Support after rescaling. However, anchoring vignettes are not a cure-all in solving measurement bias in cross-cultural surveys; we discuss the technicality and directions for further research on this technique.

Session Thu 30, 15:30 - 17:00 1

30 August 2018 15:30 - 17:00

Leekstermeerzaal

Single Paper

Educational Policy and Systems, Motivational, Social and Affective Processes

Single Paper Special groups

Keywords: At-risk students, Classroom climate, Data-based decision making and data use, Effectiveness and efficiency, Emotion and affect, Experimental studies, Literacy, Motivation

Interest group: SIG 08 - Motivation and Emotion

Chairperson: Trynke Keuning, University of Twente, Netherlands

Differential effects of autonomy support on salivary alpha-amylase

Keywords: Classroom climate, Emotion and affect, Motivation, Experimental studies

Presenting Author:Vanda Sieber, University of Zurich, Switzerland; **Co-Author:**Julia Schüler, University of Konstanz, Germany; **Co-Author:**Mirko Wegner, Humboldt-University Berlin, Germany

Several situations in learning environments can lead to psychosocial stress, which is known to be reflected in physiological processes and can constitute a serious health risk (Dickerson & Kemeny, 2004). Previous studies have shown that autonomy supportive behavior of teachers reduces physiological stress in students (Reeve & Tseng, 2011). Referring to the matching hypothesis (e.g. Schüler & Brandstätter, 2013) the present experiment tested whether an implicit autonomy disposition, which is defined as a preference for origin experiences, moderates the effect of different physical education teaching styles (controlling, autonomy supporting and neutral) on the acute physiological stress response (salivary alpha-amylase) in adolescents (N = 69). The study revealed that participants with a high implicit autonomy disposition displayed lower salivary alpha-amylase responses when exposed to autonomy-supportive vignettes compared to when they were exposed to controlling or neutral teaching styles. The opposite pattern was found in students with a low implicit autonomy disposition. The results illustrate that experimentally induced variations in autonomy support lead to different physiological stress responses, depending on individual differences in the implicit autonomy disposition.

Is part-time special education beneficial for children at risk for reading difficulties?

Keywords: At-risk students, Data-based decision making and data use, Effectiveness and efficiency, Literacy
Presenting Author:Valdar Tammik, Tallinn University, Estonia; **Co-Author:**Piret Soodla, University of Tallinn, Estonia; **Co-Author:**Eve Kikas, Tallinn University, Estonia

Effectiveness of educational support systems is often assessed by indirect means through comparative data such as that from Program for International Student Assessments (PISA; e.g., Ise et al. 2011). This longitudinal quasi-experimental study aimed to directly assess the effectiveness of part-time special education support for reading problems in Estonia, a country that has consistently ranked in the very top in PISA results despite having a support system not characterized by traits usually considered effective in the literature. Data on pre-reading and reading skills (phoneme awareness, rapid automatized naming, word reading accuracy, reading fluency) was collected from 474 first grade pupils, 70 of whom received curriculum-aligned extra lessons from special education teachers (Treatment Group). Statistical matching using a genetic search algorithm was used to find a control group that was similar to the Treatment Group on the assessed skills and parental educational level but did not receive extra help (Control Group). Students' reading fluency was again assessed at the end of grades 1 and 2 with the results indicating no statistically significant differences between the Treatment and Control groups. These findings call into question the effectiveness of a common educational support pattern and highlight the importance of *directly* assessing efficacy of support systems in contrast to using indirect retrospective evidence.

Session Thu 30, 15:30 - 17:00 2

30 August 2018 15:30 - 17:00

Conference room 2

Single Paper

Assessment and Evaluation, Educational Policy and Systems

Single Paper Equity 1

Keywords: Early childhood education, Educational policy, Effectiveness and efficiency, Equity, Large scale assessment/large scale survey, Longitudinal analyses/research, Primary education, Quantitative methods, School choice, Secondary data analysis, Student background

Interest group: SIG 18 - Educational Effectiveness

Chairperson: Hartmut Esser, Mannheim Centre for European Social Research (MZES), Germany

First Year at School in Brazil: assessment of children development and policy impact

Keywords: Educational policy, Early childhood education, Large scale assessment/large scale survey, Longitudinal analyses/research

Presenting Author:Mariane Campelo Koslinski, Federal University of Rio de Janeiro, Brazil; **Co-Author:**Tiago Lisboa Bartholo, Federal University of Rio de Janeiro, Brazil; **Co-Author:**Marcio da Costa, Federal University of Rio de Janeiro, Brazil

The study analyzes children's development in their first year at school in Rio de Janeiro using iPIPS (International Study of Children Starting School) and assess the impact of a early education policy in the city of Rio de Janeiro. The data used in the analysis was based on the first and second waves of a longitudinal study, Baseline Brazil (2017-2019). Key research questions are: 1- How much progress is made in the first year? 2- How do the children in Rio de Janeiro compare with learners in England and Scotland? 3- What is the impact of attending *Child Development Center Preschools* – main policy for early childhood education – oppose to regular preschools? A representative stratified random sample of 47 Municipal Public Schools in Rio de Janeiro (2758 children) were selected to participate in the study. Two analyses were conducted using the Rasch measurement, Winstep Software: a) Items analyses for language and mathematics; b) Distribution of the items and students on the. Two-level hierarchical linear models were used to estimate the impact of attending *Child Development Center Preschools* – main policy for early childhood education in the city. Preliminary analyses indicate that: a) the items on both tests present adequate behavior, suggesting theoretical congruence of the items, good adaptation and application protocol; b) Starting points (baseline) and gains in the first year of school are much different comparing learners from Rio de Janeiro, England and Scotland; c) data suggests a small positive impact of Child Development Centers Preschools .13 effect size for language development

Understanding the Dynamics of the Urban-Rural School Educational Gap in Chile

Keywords: Effectiveness and efficiency, Primary education, Quantitative methods, Secondary data analysis

Presenting Author:Valentina Giaconi, DEMRE, Universidad de Chile / ICED, Universidad de O'Higgins, Chile; **Co-Author:**Daniela Jiménez, DEMRE, Universidad de Chile, Chile; **Co-Author:**María Leonor Varas, DEMRE, Universidad de Chile, Chile

The urban-rural educational gap is usually associated with lower results for rural students, explained by lower socioeconomic conditions, isolation and fewer opportunities. In addition, the human development has changed the rural reality all over the world, this implies that we need new understandings of the rural-urban educational gap. In the case of Chile, several authors have raised the importance of deepen the comprehension of rural education because it has systematic lower results (Raczynski & Román, 2016). In this proposal, we estimate school-level gaps in standardized test

scores between urban and rural schools. We used the scores from a national standardized test applied to 4th graders during the years 2002 to 2016. In the raw comparison, rural schools have consistently worst results than non-rural schools. However, if we select only public schools, we found that the gap changes direction. This trend is explained with decreasing inequalities regarding the socioeconomic composition of urban and rural schools, but also rural schools appear to be more effective than urban schools.

Inequalities and Educational Decisions: Role of Pupils' Cultural Capital in Secondary School Choice

Keywords: Equity, School choice, Student background, Primary education

Presenting Author:Iva Kosutic, Institute for Social Research - Zagreb, Croatia; **Co-Author:**Jelena Matić, Institute for Social Research - Zagreb, Croatia, Croatia; **Co-Author:**Boris Jokić, Institute for Social Research - Zagreb, Croatia

This paper examines social inequalities in the educational decision-making process with a sample of elementary school pupils in the City of Zagreb, Croatia (N=1031). The relationship between pupils' social background, educational achievement and their educational choices, representing important components of life choices and outcomes in general, is a frequently addressed topic in research. The quality and effectiveness of a school is largely dependent on the ability of the school to reduce inequalities in educational opportunities for children from diverse economic and educational backgrounds. The theoretical framework used in this paper is based on Bourdieu's cultural and social reproduction theory (1977, 1984). Consistent with Bourdieu's assumption that pupils with higher levels of cultural capital hold higher educational aspirations and favour more prestigious types of education (gymnasiums), this paper examines the role of the cultural capital of elementary school pupils in predicting their choice of secondary school (vocational education or gymnasium). Pupils' cultural capital was operationalised as cultural practices (present and past) and the mother's educational level. Logistic regression was used for predicting the secondary school choice of pupils in their final year of elementary education (8th grade). Analyses indicated that cultural capital indicators were statistically significant predictors for the choice of secondary school, in line with the theoretical assumption. The results are discussed in light of possible strategies for decreasing the effects of pupils' cultural capital and social background on their educational chances and diminishing social inequalities that could be maintained through educational inequalities.

Session Thu 30, 15:30 - 17:00 3

30 August 2018 15:30 - 17:00

Auditorium

Symposium

Assessment and Evaluation

Factors influencing data use in schools

Keywords: Accountability and school inspections, Data-based decision making and data use, Educational policy, Improvement, Large scale assessment/large scale survey, Mixed-method research

Interest group: SIG 23 - Educational Evaluation, Accountability and School Improvement

Chairperson: Tobias Feldhoff, Johannes Gutenberg University of Mainz, Germany

Organiser: Sebastian Wurster, Johannes Gutenberg University Mainz, Germany

Discussant: Melanie Ehren, United Kingdom

Data-based decision making plays a key role for governmental strategies of school improvement in Europe. In most European countries self-evaluation procedures, school inspections and feedback systems of student achievement data were implemented. Policy expects that these data sources are used for improvement activities. Despite the high relevance of data use for adequate improvement measures, providing data for schools does not automatically result in improvement. Different studies identified various factors influencing data use. However, the presence of multiple data sources in a multilevel system with a complex configuration of enablers and barriers of data use, still requires further research on crucial processes of data use and factors influencing these. The symposium addresses different key issues of data use regarding different sources which haven't been studied yet: 1) the validity and data quality of self-evaluations which are important preconditions of initiating appropriate improvement measures, 2) an extended analytical differentiation of the steps in data use processes and the question if factors influencing teacher's data use differ between multiple data sources and 3) the functioning of state-wide support systems and school supervising authority, their influence on data use and the different roles of the multiple parties involved in the support system.

Support systems for initiating school development improvement through school inspection results

Presenting Author:Tanja Rettinger, Johannes Gutenberg University Mainz, Germany; **Co-Author:**Sebastian Wurster, Johannes Gutenberg University Mainz, Germany; **Co-Author:**Tobias Feldhoff, Johannes Gutenberg University of Mainz, Germany

Data-based school improvement is an essential goal of current efforts to ensure and improve school quality in Germany (Altrichter & Maag Merki 2016). Within the last 10 years, school inspections have become an important instrument to measure school quality and were implemented for two reasons: to improve and to control school quality. To what extent results from school inspections are used for improvement measures is therefore a central question. There are individual as well as organizational factors of interest, which (can) promote or hinder data use in schools. However, little is known about

support systems regarding their role in promoting school improvement based on school inspections. The aim of this study is to provide empirical insight about the importance of support systems as a relevant factor for data use of inspection results.

Using school self-evaluation results for school development. What about validity and data quality?

Presenting Author: Jerich Faddar, University of Antwerp, Belgium; **Co-Author:** Sven De Maeyer, University of Antwerp, Belgium; **Co-Author:** Jan Vanhoof, University of Antwerp, Belgium

Factors fostering and hindering teacher's use of multiple data sources. A large scale study.

Presenting Author: Sebastian Wurster, Johannes Gutenberg University Mainz, Germany

Session Thu 30, 15:30 - 17:00 4

30 August 2018 15:30 - 17:00

Conference room 1

Symposium

The implementation of Success for All in the Netherlands, highlighted from different perspectives

Keywords: At-risk students, Collaborative Learning, Content analysis, Educational reform, Effective instruction and teaching, Literacy, Mixed-method research, Parental involvement in learning, Primary education, Program implementation and evaluation, Teacher professional development

Interest group:

Chairperson: Emmeline van der Scheer, University of Groningen, Netherlands

Discussant: Robert Slavin, Johns Hopkins University, United States

In this symposium, the requirements, strengths and challenges of implementing the whole-school reform program Success for All (SfA) in the Netherlands will be discussed at the level of the school, class, students and parents. The audience is provided with an overview of the Dutch SfA program at the start of the symposium. The first research paper discusses the results from a review study focusing on the implementation of SfA in the US and UK context, and connecting these results to the status quo of the Dutch context. In this contribution, the focus will be on the entire program and the level of the school and the class will be central. In the subsequent papers, the focus will be on specific components or mechanisms through which SfA is hypothesized to attain its positive results in student attainment. The second paper specifically addresses the implementation of cooperative learning – one of the key elements in SfA's instructional program - in the Dutch classrooms. The third paper describes the relationship between students' task-oriented behavior in the classroom and SfA NL. The last paper highlights the importance of parental involvement in reading by discussing the results of a systematic review, drawing implications for SfA NL's home involvement-component.

Implementing Success for All in an international context: results of a systematic review

Presenting Author: Marijke Mullender-Wijnsma, Rijksuniversiteit Groningen, Netherlands; **Co-Author:** Mechteld van Kuijk, University of Groningen, Netherlands

While many studies have been conducted on the effectiveness of Success for All (SfA), only limited studies have focused on the implementation of this whole-school reform program. This is considered a remarkable finding, as multiple studies that evaluated the effects of SfA indicated that these effects would have been stronger if the program would have been implemented to a higher degree. Different implementation studies address different components of SfA, resulting in the need for an overview of these studies' results in relation to a components-framework. The results will be compared to those aspects known to be relevant for effective implementation of programs. As such, it can be identified to what extent 'dose delivered' or 'dose received' have been affected. The results show that in addition to a small number of pertinent 'implementation struggles' (such as the prescribed nature of the instructional program), there is a large diversity in the factors attributing to success or failure of the program. As part of the conference symposium, connections will be made to the current implementation status of several components in the context of 'SfA the Netherlands'.

Implementation of cooperative learning: differences in teachers' attitudes and beliefs.

Presenting Author: Marij Veldman, University of Groningen, Netherlands; **Co-Author:** Mechteld van Kuijk, University of Groningen, Netherlands; **Co-Author:** Simone Doolaard, University of Groningen, Netherlands; **Co-Author:** Roel J. Bosker, Rijksuniversiteit Groningen, Dept of Education and GION, Netherlands

Levels of implementation of cooperative learning (CL) in teachers' day-to-day practices in the context of Success for All lessons in the Netherlands differed strongly between teachers, especially in teaching students the needed cooperative behaviors by modelling good interaction skills and reflecting on, encouraging, and evaluating group work. Using video observations, high- and low-performing CL teachers were distinguished and, subsequently, interviewed to gain insight into their attitudes and beliefs. Teachers who differed in their levels of implementation of cooperative learning also differed in their attitudes and beliefs about the approach. High-performing CL teachers experienced positive changes in students' behavior and learning outcomes, while low-performing CL teachers appeared to be less convinced of the value of the strategy and struggled more with children's behavior during cooperative learning. It can be concluded that implementing cooperative learning is highly complex and supporting teachers, also during and after implementation of their new

practices, is essential.

The relationship between students' task-oriented behavior and Success for All

Presenting Author: Mariëtte Hingstman, Rijksuniversiteit Groningen, Dept of Education and GION, Netherlands; **Co-Author:** Simone Doolaard, University of Groningen, Netherlands; **Co-Author:** Roel J. Bosker, Rijksuniversiteit Groningen, Dept of Education and GION, Netherlands

The objective of this study is to describe the relationship between Success for All and task-oriented behavior of students with attention difficulties. Attention difficulties appear to be a strong predictor for poor academic achievement. Individual characteristics like gender, age, motivation and socioeconomic background affect children's attention processes, although engagement in academic activities is also dependent on contextual factors, like the classroom climate. Therefore, it should be a high priority to examine how students' engagement can be stimulated. Several components of the Success for All (SfA) program are expected to increase task-oriented behavior. Through systematic direct observations with a time-sampling observation instrument, it is investigated if SfA affects the behavior of students with attention difficulties. In addition, the relationships of task-oriented behavior with setting and with teacher activity were investigated. The design is quasi-experimental. 72 first-grade students from 10 primary schools participated in the study. The experimental condition (SfA) contained 10 classes (40 students), the control condition contained 8 classes (32 students). In every classroom 4 students are observed during a full reading lesson. Those were the students with the lowest score (compared to their classmates) on the subscale 'Task-performing' of the SCOL (Social Competence Observation List) questionnaire. A multilevel logistic regression model is used for the analysis.

Spontaneous and School-initiated Home-based Parental Involvement in Reading in grade 1- 3.

Presenting Author: Marlot Akkermans, University of Groningen, Netherlands

Parental involvement is a key concept in children's educational success. Especially home-based parental involvement in reading offered to children in the first Grades of elementary school has the potential to influence children's reading achievement. In this review we explore 19 recently published (2000 – 2017) peer-reviewed observational, questionnaire, and intervention studies on both spontaneous and school-initiated parental involvement and its association with the reading proficiency of first, second, and third Grade children. The case is made that parental involvement in beginning reading has the potential to benefit children's reading inclination and/or proficiency, but that such an outcome requires a large amount of willingness and investment of time and energy of both parents and educators. With respect to spontaneously occurring home-based early literacy involvement, it appears to be the Home Literacy Environment (HLE) and the quality of the parent-child interaction that are most strongly related to children's inclination to read and their reading skills. School-initiated interventions designed to enhance the quality and quantity of parental involvement in beginning reading are characterized by an emphasis on the training of parents in school-like literacy practices. Researchers seem to disagree on whether parents should focus on comprehension, technical reading skills, or both.

Session Fri 31, 09:00 - 10:30 1

31 August 2018 09:00 - 10:30

Conference room 2

Single Paper

Assessment and Evaluation, Educational Policy and Systems

Single Paper Equity 2

Keywords: Achievement, Class and school composition, Cross-national/cultural comparisons/International comparisons, Equity, Ethnography, Hierarchical linear modeling, Large scale assessment/large scale survey, Quantitative methods, Secondary data analysis, Student background

Interest group: SIG 18 - Educational Effectiveness, SIG 23 - Educational Evaluation, Accountability and School Improvement

Chairperson: Mariëtte Hingstman, Rijksuniversiteit Groningen, Dept of Education and GION, Netherlands

The relation between school composition and science performance in TIMSS 2015

Keywords: Achievement, Class and school composition, Student background, Hierarchical linear modeling

Presenting Author: Mario Vennemann, University of Paderborn, Germany; **Co-Author:** Birgit Eickelmann, Paderborn University, Germany

This contribution addresses the question of whether or not there are effects of the school body composition towards students with certain characteristics that showed to be confounded with substantial educational disparities (such as gender, migratory background, low cultural capital and low socio-economic status) from the perspective of School Effectiveness Research (SER). In doing so, this paper focuses on the science achievement of primary school students from the Trends in International Mathematics and Science Study (cf. Martin, Mullis, Foy, & Hooper, 2016) and utilizes hierarchical linear modelling (HLM) techniques (cf. Raudenbush & Bryk, 2002) to examine the relevance of the achievement-related and social school body composition in three European educational systems (Flemish community in Belgium, Denmark, Germany) comprising a student sample of about N = 13.000 students from fourth grade. The results of a 3-step analytical approach indicate that compositional measures of schools are relevant in all countries under research and that there are

country-specific differences in the relevance of the achievement-related and social student body composition in the science domain. Additionally, comparing these results of this paper to those for the mathematics domain raises the question of why the effect of school composition seems to be domain specific.

Trends in social and ethnic inequity in the three Belgian Communities

Keywords: Equity, Cross-national/cultural comparisons/International comparisons, Ethnography, Large scale assessment/large scale survey

Presenting Author:Valérie Quittre, Research unit EQUALE, University of Liege, Belgium; **Co-Author:**Françoise Crépin, University of Liège, Research Unit EQUALE, Belgium; **Co-Author:**Dominique Lafontaine, University of Liège, Research Unit EQUALE, Belgium

Using PISA data from 2006 to 2015, this study investigated the trend in social and ethnic inequity in the Belgian communities. The three educational systems are highly stratified and are among the most inequitable both for socio-economically disadvantaged and for immigrant students. However, the three Belgian communities are facing dissimilar contexts and issues regarding the migrant population denoting that inequity issues require investigation at the region level. The study showed that in the French and the Flemish speaking communities, the interaction between socio-economic and immigration background is significant meaning that on average the immigrant students do not really benefit the contribution of a higher socioeconomic status to the same extent the native students do. This effect is higher in Flanders showing that social background and migration background interact differently in the two regions. Moreover, through time –from 2006 to 2015 -, in the French speaking community the gap of achievement between native and non-native students after control of the socio-economic status has narrowed down. This is not the case in Flanders. Thus, it appears that the systemic discrimination acts to a larger extent in the Dutch part compared to the French part. The migration issue in the German speaking community is analysed separately since this border area shows an atypical pattern and welcomes number of German students with a socio-economic level higher than those of the native students.

Decomposing the Ethnic Achievement Gap in Colombia

Keywords: Achievement, Equity, Quantitative methods, Secondary data analysis

Presenting Author:Beatriz Gallo Cordoba, University of Bristol, United Kingdom

Achievement gaps are the results of processes at different levels of the education system. Identifying the contribution to the achievement gap of each level is crucial for research and policy efforts that result in sustainable school improvement. Nonetheless, the recent debate on how to decompose achievement gaps into its within- and between-school gaps is somewhat unresolved. This paper argues that the debate is based on the confusion between measuring the gaps at each level and estimating their contribution to the overall gap. The paper adds a level of analysis by decomposing the ethnic achievement gap of Colombia into its within-school, between-school and between- Local Authority components. Using the SABER 11 dataset, the distinction is made between the achievement gaps at each level and the contribution of each level to the overall achievement gap. Results contrast to those reported in the context of the debate in the US as within-school gaps are the smallest contributors to the overall achievement gap. In turn, differences between Local Authorities contribute the most to the overall gap, although the contribution of differences between schools is not negligible. A call is made for future research and rethinking policy.

Session Fri 31, 09:00 - 10:30 2

31 August 2018 09:00 - 10:30

Leekstermeerzaal

Single Paper

Educational Policy and Systems, Motivational, Social and Affective Processes

Single Paper Tracking

Keywords: Ability grouping and tracking, Achievement, Class and school composition, Developmental processes, Educational attainment, Motivation, Quasi-experimental research, Secondary education, Student background, Vocational education

Interest group: SIG 18 - Educational Effectiveness

Chairperson: Marij Veldman, University of Groningen, Netherlands

Factors affecting the educational and occupational aspirations of students in a tracked system.

Keywords: Ability grouping and tracking, Motivation, Student background, Secondary education

Presenting Author:Jana Strakova, Faculty of Education, Charels University in Prague, Czech Republic; **Co-Author:**Jaroslava Simonova, Faculty of Education, Charels University in Prague, Czech Republic

The aim of the paper is to contribute to the exploration of educational inequalities in the Czech education system by studying the educational and occupational aspirations of students at the end of their compulsory education. After completing compulsory education students can apply for three upper secondary tracks with completely different curricula: an academic track and two vocational tracks: technical and apprenticeship. The decision about the application stays with the students and their families; the school does not participate in the decision-making process. This study intends to show that the current system is disadvantageous for students from low social strata because they are forced to choose vocational

education, which limits their future opportunities. It explores the extent to which the choice of upper secondary education and ideas about future occupation are predicted by academic achievement and school grades and the extent to which they are predicted by socio-economic background. It also explores the concept of academic futility and its relationship to the academic and occupational aspirations of young Czechs. The analyses are carried out on data from 4237 grade 9 students collected in 2016 within the Czech Longitudinal Study in Education and data from the Czech PISA-L 2003 study, with 6340 cases. The main analytical method used in the analyses is logistic regression.

Is ability tracking really responsible for educational inequality?

Keywords: Ability grouping and tracking, Achievement, Class and school composition, Educational attainment

Presenting Author:Hartmut Esser, Mannheim Centre for European Social Research (MZES), Germany

The contribution addresses the question of whether ability tracking really reinforces educational inequalities without improving achievement as it is stated in most contributions to the topic as standard position. Starting point is a general theoretical model for identifying the effects of ability tracking over the course of educational careers from family, preschool, elementary school, transition to secondary school and later achievements. Because some of the relevant conditions of the model are not available in international studies (like PISA), effects for the (rather large) variation in institutional regulations of the tracking systems of the 16 German country states are analyzed by data of the "National Educational Panel Study" (NEPS), which contains all relevant conditions. Empirical results show a clear picture: (Strict) Ability Tracking does not reinforce educational inequalities by social background, but increases "meritocratic" sorting by strengthening the effect of cognitive ability resp. prior achievements for attainment, enlarges cognitive homogenization of school classes and amplifies the (in general positive) effects of ability composition of schools/school classes on achievements. Main conclusion for improving school-effectiveness (without negative consequences on social educational inequalities) would be: A *more* stringent implementation of differentiation by ability and prior achievement by improving reliability and validity of teacher's evaluations and limiting parent's decisions on attainment against recommendations.

Track effects on academic self-concepts

Keywords: Ability grouping and tracking, Developmental processes, Vocational education, Quasi-experimental research

Presenting Author:Bieke De Fraine, KU LEUVEN, Belgium; **Co-Author:**Jonas Dockx, KU LEUVEN, Belgium; **Co-Author:**Machteld Vandecandelaere, KU Leuven, Belgium

This study investigated the development in general academic self-concept and domain-specific self-concepts for mathematics and Dutch in four different tracks. A longitudinal cohort study in Flanders (3205 students in 46 schools) was used to compare students' academic self-concepts during the first three years of secondary education. The investigated tracks differ in the extent their educational programs are more academically or more vocationally focused and differ in average student academic ability. To control for selection effects, students who are comparable across the four tracks were matched using propensity score matching, Mahalanobis distance matching and coarsened exact matching. By means of multiple indicator quadratic latent growth curves, pairs of tracks that are hierarchically consecutive were compared regarding the development in academic self-concepts. For the two highest tracks, it was beneficial to be allocated to the highest track, whereas the pairwise comparisons between the three lower tracks indicated a detrimental effect of being in a higher track. Effects generally only becoming apparent after two years.

Session Fri 31, 09:00 - 10:30 3

31 August 2018 09:00 - 10:30

Conference room 1

Single Paper

Assessment and Evaluation, Educational Policy and Systems, Teaching and Teacher Education

Single Paper Teachers

Keywords: Accountability and school inspections, Data-based decision making and data use, Effective instruction and teaching, Effectiveness and efficiency, Experimental studies, Improvement, Primary education, Teacher professional development

Interest group: SIG 18 - Educational Effectiveness, SIG 23 - Educational Evaluation, Accountability and School Improvement

Chairperson: Simone Doolaard, University of Groningen, Netherlands

Direct and Indirect Effects of the Dynamic Approach to School Improvement on Learning Outcomes

Keywords: Effectiveness and efficiency, Improvement, Primary education, Experimental studies

Presenting Author:Leonidas Kyriakides, University of Cyprus, Cyprus; **Co-Author:**Bert P.M. Creemers, University of Groningen, Netherlands; **Co-Author:**Evi Charalambous, University of Cyprus, Cyprus

Drawing on research on the dynamic model of educational effectiveness a dynamic approach to promote quality in schools has been developed. The main assumptions of the dynamic approach to school improvement (DASI) are presented and its main steps are described. This study investigates the extent to which DASI can help schools in socially disadvantaged areas to develop strategies and action plans in order to improve their learning environment and their policy for teaching and through that to promote student learning outcomes in Mathematics. Specifically, at the beginning of the school year 2015-

2016, a sample of 72 primary schools in four European countries (Cyprus, England, Greece and Ireland) was selected and randomly split into the experimental and control groups. We administered a questionnaire measuring the school factors of the dynamic model (i.e. learning environment, their policy for teaching, and their evaluation mechanisms) to all teachers of the school sample (n=762) and a battery of Mathematics tests to students of Grades 4-6 (n=5560). Using multilevel SEM techniques, both direct and indirect effects (through improving the school factors) of the intervention on promoting student learning outcomes were identified. However, the direct effect of DASI was found to be much stronger than its indirect effect. Implications for research, policy and practice are, finally, drawn.

Capturing the Complexity of Differentiated Instruction

Keywords: Data-based decision making and data use, Effective instruction and teaching, Teacher professional development, Primary education

Presenting Author: Trynke Keuning, University of Twente, Netherlands; **Co-Author:** Marieke van Geel, University of Twente, Netherlands

Providing differentiated instruction (DI) is regarded an important but complex teaching skill, which many teachers do not master or feel prepared for. In order to design professional development activities, a thorough description of this task is required. Various instruments can be found in the international literature for assessing teachers' differentiation qualities, from self-reports to observation schemes, and from perceived-difficulty instruments to student questionnaires. The question is whether these instruments truly capture the complexity of differentiation. In order to depict this complexity, a Cognitive Task Analysis (CTA) of the differentiation skill was performed. The resulting differentiation skill hierarchy is presented here, together with the knowledge required for differentiation, and the factors influencing complexity. Furthermore, six underlying 'principles of differentiation' were identified. Based on the insights of this CTA, professional development trajectories can be designed, and a comprehensive assessment instrument can be developed, enabling researchers and practitioners to train, assess and monitor teaching quality in terms of providing differentiated instruction.

Cognitive Activation and Mathematical Richness: Predicting Different Teaching Profiles

Keywords: Accountability and school inspections, Effective instruction and teaching, Effectiveness and efficiency, Primary education

Presenting Author: Evridiki Kasapi, University of Cyprus, Cyprus; **Presenting Author:** Charalambos Charalambous, University of Cyprus, Cyprus

Instruction represents a complex system of interactions among the teacher, the students, and the content. Given the complexity of this system, several theoretical and conceptual frameworks have been developed to capture instructional quality and through that understand its contribution to student learning. In mathematics, the subject-matter of interest in this paper, scholars have attended to issues of *Cognitive Activation* and *Mathematical Richness* as two key contributing teaching factors to student learning. The cross-tabulation of these factors can lead to four distinct teaching profiles, depending on the combinations of these two features at high or low levels. Scholarly work around this framework is, however, in very early stages; empirical research on whether these teaching profiles are likely to vary according to teacher and lesson characteristics is still scarce. To address this gap, we analyzed 228 videotaped lessons from a convenient sample of 38 teachers (each giving six lessons), using the *Mathematical Quality of Instruction* instrument; the teachers' *Mathematical Knowledge for Teaching* (MKT) was also measured. Using Latent Profile Analysis, we investigated whether the aforementioned four teaching profiles were empirically validated and whether specific teacher and lesson characteristics could predict the distribution of each teacher's lessons in the profiles. The results suggested a solution of five profiles that was not, however, too distant from the profiles theorized; teachers' MKT and lesson content were key factors found to contribute to classification in these profiles. We discuss the implications of this work for teacher and teaching evaluation and for teacher professional development.

Session Fri 31, 09:00 - 10:30 4

31 August 2018 09:00 - 10:30

Auditorium

Symposium

Assessment and Evaluation, Educational Policy and Systems

Malleability in school and system effectiveness; what are realistic expectations about effect sizes?

Keywords: Achievement, Comparative studies, Cross-national/cultural comparisons/International comparisons, Educational reform, Effectiveness and efficiency, Improvement, Large scale assessment/large scale survey, Large scale assessment/large scale survey, Longitudinal analyses/research, Primary education, Quantitative methods

Interest group: SIG 18 - Educational Effectiveness, SIG 23 - Educational Evaluation, Accountability and School Improvement

Chairperson: Jaap Scheerens, University of Twente, Netherlands

Organiser: Jaap Scheerens, University of Twente, Netherlands

Discussant: Hans Luyten, University of Twente, Netherlands

A key issue in educational effectiveness research is separating the effects of malleable, policy amenable factors on student achievement from "given" background conditions at system and student level. Recent studies indicate that the effects of

background conditions tend to be bigger, and those of malleable variables and interventions smaller, than usually expected (Marks, 2015, Scheerens, 2016). Comparative studies show that, generally, the performance of educational systems is quite stable across time, and that reforms take long to have effects, if they have effects at all. This combination of findings gives reason to pose “limited malleability”, as an emerging perspective in educational effectiveness theory and research. The objective of this symposium is to further explore the thesis of “limited malleability” in three empirical studies, and to try and make up the balance on realistic effect sizes. The first study used a combination of internationally comparative datasets to explore the correlations of malleable and non-malleable characteristics of structural and culture characteristics of educational systems with student achievement. The second study was based on secondary analyses of five subsequent waves of the PISA study, and analyzed the stability of country performance over time and the influence of curriculum innovations. The third paper analyzed differences in effectiveness between schools by using the learning gain of students over three points in time by means of growth curve modelling and used data from the pupil monitoring system in Dutch primary education. The results partially support the thesis of limited malleability, but also show interesting nuances and challenges.

Country-Level Correlates of Educational Achievement: Evidence from Large-Scale Surveys

Presenting Author:Jia He, TilbuDIPF, China; **Co-Author:**Fons van de Vijver, Tilburg University, Netherlands; **Co-Author:**Alena Kulikova, National Research University Higher School of Economics, Moscow, Russian Federation

Linking country-level educational achievement with other country-level indicators has the potential to drive systemic educational changes, as these correlates may reflect characteristics relevant for policy-making decisions to improve educational effectiveness. This study establishes a nomological network of educational achievement at country level, with clusters of country variables derived from psychological, sociological, and other country-comparative research. Country educational achievements were compiled from all cycles of the Program for International Students Achievement, and the Trends in International Mathematics and Science Study for Grade 8 and Grade 4 students. These scores were found to be correlated with (1) affluence and educational investment indicators (positive), (2) intelligence (positive), (3) religious diversity (positive), (4) values pertaining to long-term orientation (positive), autonomy (positive), uncertainty avoidance (negative) and embeddedness (negative), and (5) aggregated scores of conscientiousness, openness, self-esteem, and teacher self-reports of job satisfaction and efficacy (negative). Policy implications are discussed.

The modest impact of education reforms on PISA trends

Presenting Author:Peter Tymms, University of Durham, United Kingdom; **Co-Author:**Cesare Aloisi, Centre for Education Research and Practice, United Kingdom

The stability of educational test results from PISA over 15 years was examined and the influence of demographics and social capital was assessed, as was the impact of educational reforms. The test results were remarkably stable, with correlations up to 0.99 for country-level results over two cycles. Despite this stability, trends were observed with scores generally rising year on year, but so too were the socio-economic indicators. Together with measures of gender, immigration, OECD membership and first language, these variables went a long way to account for the rising scores. Case studies suggest that the clearest impact of reforms on test scores amount to an annual Effect Size of around 0.02. The paper argues for the greater prominence of fairer adjusted PISA league tables and suggests that multi-disciplinary approaches to educational data analyses and policy advice are needed.

Growth curve analyses to estimate the size, stability and consistency of school effects

Presenting Author:Anneke Timmermans, University of Groningen, Netherlands; **Co-Author:**Greetje van der Werf, University of Groningen, Netherlands

The current study explored the size, stability, and consistency of school effects, using two different effectiveness indicators, being the achievements of students at the end of primary school (gross school effects) and the growth in achievement across three years of schooling (value added effects). The empirical data are from a Dutch data set including the scores of 25,269 students on Reading comprehension, Spelling, and Mathematics tests, taken in grade 4, 5 and 6 among three cohorts of students in 319 primary schools. Multivariate multilevel growth curve modelling was applied in which the measurements (level 1) were nested within students (level 2), which were nested in school-cohorts (level 3), which were nested in schools (level 4). The results showed that 1) for students' growth of achievement the relative proportion of variance at the school level seemed larger compared to achievement at the end of primary school; 2) the total variance in growth was substantially smaller compared to variance in achievement at the end of primary school; 3) school effects for growth are less stable across different cohorts of students than school effects established at a particular point in time; and 4) school effects for growth over time are less consistent across multiple subject domains than school effects as indicated by students' achievement at a particular moment. Differences with findings from previous research are discussed.

Session Fri 31, 11:00 - 12:30 1

31 August 2018 11:00 - 12:30

Conference room 1

Single Paper

Educational Policy and Systems

Single Paper Accountability

Keywords: Accountability and school inspections, Cross-national/cultural comparisons/International comparisons, Educational policy, Effectiveness and efficiency, Emotion and affect, Hierarchical linear modeling, Large scale assessment/large scale survey, Qualitative methods, Social aspects of learning and teaching, Teacher professional development

Interest group: SIG 23 - Educational Evaluation, Accountability and School Improvement

Chairperson: Tina Montreuil, McGill University, Canada

Trust and accountability to improve learning outcomes; a systematic literature review

Keywords: Accountability and school inspections, Educational policy, Effectiveness and efficiency, Qualitative methods

Presenting Author: Melanie Ehren, UCL Institute of Education, United Kingdom; **Co-Author:** Jacqueline Baxter, The Open University, United Kingdom; **Co-Author:** Andrew Paterson, JET, South Africa

Accountability and trust are key variables in the improvement of any education system and are often positioned as opposites or substitutes, saying that control destroys trust or that control is unnecessary when you trust someone. Others however argue that control can build trust when it for example locks people or organizations into a collaborative relationship. The choice to trust and cooperate and engage in school accountability is also bound by the institutional context in which people interact, such as the political, legal and economic framework, and informal rules, socially accepted norms and patterns of behaviour in a country (Zaheer et al. (1998; Lyon et al, 2015, p.7). These institutions influence how people interact both consciously and unconsciously; they provide meaning to the circumstances before a relationship is built and they influence the patterns of how people interact when they start to actively establish a relationship (Van der Voort (2017). Interpersonal relations of accountability and trust are thus embedded in an institutional context, both temporally, socially, and institutionally. This paper presents the outcomes of a systematic literature review to understand how accountability and trust are related (both on the interpersonal and system level) to improve learning outcomes over time.

Teacher accountability policy, sociocultural context, and student outcomes across countries

Keywords: Accountability and school inspections, Cross-national/cultural comparisons/International comparisons, Hierarchical linear modeling, Large scale assessment/large scale survey

Presenting Author: Yue-Yi Hwa, University of Cambridge, United Kingdom

How does national sociocultural context influence the relationships between teacher accountability instruments, teacher motivation, and student outcomes, as measured in cross-country surveys? While there is a growing recognition of the need for teacher accountability policies to be compatible with local contexts, existing cross-country statistical analyses of the effect of teacher accountability on student outcomes do not take sociocultural context into account. In this study, I use multilevel statistical modelling to examine the interaction between teacher accountability instruments and national sociocultural context. I test a series of cross-sectional models using data on student outcomes and teacher motivation from PISA 2012, PISA 2015, and TIMSS 2015; measures of formal teacher accountability from PISA 2012 and 2015; and a range of national sociocultural constructs derived from the World Values Survey, European Values Study, and Hofstede's IBM dataset. The models indicate that certain sociocultural constructs (specifically, social trust and confidence in institutions) are associated with a stronger relationship between teacher accountability instruments and student outcomes, affecting both the level and equity of student achievement. Additionally, a different set of sociocultural constructs (i.e. power distance and uncertainty avoidance) moderate the relationship between teacher accountability instruments and teacher motivation. In demonstrating that sociocultural context influences the efficacy of teacher accountability instruments in facilitating desirable student outcomes, this analysis challenges certain decontextualized strands of the 'best practices' education policy discourse. It also lays the groundwork for a subsequent qualitative study of the mechanisms and contexts involved in effective teacher accountability policy in two high-performing education systems.

Cognitive and Affective Responses to School Inspection Feedback

Keywords: Accountability and school inspections, Emotion and affect, Social aspects of learning and teaching, Teacher professional development

Presenting Author: Amy Quintelier, University of Antwerp, Belgium; **Co-Author:** Jan Vanhoof, Antwerp University, Belgium; **Co-Author:** Sven De Maeyer, Antwerp University, Belgium

Studies have shown that teachers are not always willing to accept the school inspection's feedback and use it for the improvement of their own teaching and learning processes. Literature distinguishes several aspects of feedback that stimulate or hinder feedback acceptance, such as recipient's cognitive and affective responses to feedback. In terms of cognitive responses, the source's perceived credibility, the perceived fairness of feedback processes and outcomes, and characteristics such as sign, constructiveness, and accuracy are distinguished. With regard to affective responses, positive affective responses to feedback increase feedback acceptance, while negative affective responses (anger, anxiety, or sadness) decrease this chance. This study investigates cognitive and affective responses of teachers during feedback reception in a school inspection context, and their influence upon feedback acceptance. The study draws on data from 21 in-depth interviews with teachers in eight primary schools. Interviews were administered until three months after the inspection. Beside the inspector's positive attitude and teachers' perceived procedural fairness, results also indicate that specific, constructive feedback is more accepted than vague, generalized feedback. Under these circumstances, emotions of joy, such as happiness and relief, are expressed. Conversely, respondents reject feedback more easily when inspectors

are perceived to be inadequately informed, arrogant, or disrespectful. When negative feedback is rated as unfair, negative emotions, that evoke during feedback reception, interfere with feedback acceptance. This study enlarges the knowledge base on the relationship between teachers' emotions and professional self-understanding in the inspection visit context and illustrates the different emotional reactions of teachers.

Session Fri 31, 11:00 - 12:30 2

31 August 2018 11:00 - 12:30

Leekstermeerzaal

Single Paper

Educational Policy and Systems, Teaching and Teacher Education

Single Paper Reading

Keywords: Cross-national/cultural comparisons/International comparisons, Educational attainment, Literacy, Longitudinal analyses/research, Primary education, Program implementation and evaluation, Quantitative methods, Teaching approaches

Interest group: SIG 18 - Educational Effectiveness, SIG 23 - Educational Evaluation, Accountability and School Improvement

Chairperson: Lionel Alvarez, Switzerland

Promoting Students Reading Literacy in an Extracurricular Activity in German All-day Primary Schools

Keywords: Literacy, Program implementation and evaluation, Primary education, Longitudinal analyses/research

Presenting Author:Karin Lossen, Technical University of Dortmund, Germany; **Co-Author:**Katja Tillmann, Technical University of Dortmund, Germany; **Co-Author:**Frederik Osadnik, Technical University of Dortmund, Germany; **Co-Author:**Heinz Günter Holtappels, Technical University of Dortmund, Germany

Enhancing students' achievement and reducing unequal educational chances are central objectives of implementing all-day schools in Germany (Aktionsrat Bildung, 2013). Based on data from the longitudinal study StEG-P: 'Study on the development of all-day schools – Quality of Extracurricular Activities and Individual Effects in All-day Primary Schools' (StEG-Konsortium, 2016) the authors firstly explore to what extent participation in reading related extracurricular learning arrangements effects the development of reading literacy (Bremerich-Vos et al., 2012). Secondly, main features of a tailor-made reading program for all-day primary schools named 'Detective-Club' are presented. It was invented in 2016 and is core of the intervention study 'StEG-Reading' (2016-2019). The paper addresses the following questions: 1.) Do students benefit from taking part in reading-related extracurricular learning arrangements with regard to their reading literacy?; 2.) Does the student rated quality of reading-related extracurricular learning arrangements influence the effects on reading competencies? 3.) Is there selectivity in students' participation in extracurricular reading activities? 4.) How could a 'tailormade' reading program for all-day primary school look like?

There are no effects of participation in reading related extracurricular activities on the development of reading literacy, even when quality of learning arrangements is rated as high. Participants as well as non-participants show an increase in reading achievement over time. The gap between students with and without migration background and from different social backgrounds is persistent. Schools that carried out the reading activity 'Detective-Club' in 2016 reported that the students' reading motivation increased and that especially the poorer readers benefit from the program.

Typologies of teaching reading in primary school – relationships with achievement and motivation

Keywords: Educational attainment, Teaching approaches, Cross-national/cultural comparisons/International comparisons, Quantitative methods

Presenting Author:Grace Grima, Pearson UK, United Kingdom; **Co-Author:**Jenny Lenkeit, Oxford University, United Kingdom; **Co-Author:**Therese N. Hopfenbeck, University of Oxford, United Kingdom; **Co-Author:**Joshua McGrane, University of Oxford, United Kingdom

Research with international large-scale assessments has neglected to investigate differences at the classroom level in order to explain variance in achievement between education systems. The classroom level is however of particular interest for policymakers, as it constitutes malleable structures and characteristics, closest to the individual level of achievement. This research uses data from the Progress in International Reading Literacy Study (PIRLS) 2016 to investigate how information on instructional practices can be mapped into 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 class analysis to describe typologies of teaching that better reflect the fact that teachers apply a combination of practices in their classroom. Further mediation analysis is conducted to evaluate the relationship between teaching typology, motivational characteristics and reading achievement within and across seven education systems participating in PIRLS 2016. Additionally, we draw on national student data from England to examine the extent to which this relationship is affected by students' prior achievement. Findings will enhance our understanding of differences in achievement between education systems and will provide an impetus for enhancing teacher qualifications at national levels.

Session Fri 31, 11:00 - 12:30 3

31 August 2018 11:00 - 12:30

Conference room 2

Single Paper

Teaching and Teacher Education

Single Paper Improvement school level

Keywords: Case studies, Culture, Curriculum change and development, Effectiveness and efficiency, Higher education, Program implementation and evaluation, Quantitative methods, Secondary education, Social interaction, Teacher professional development

Interest group: SIG 18 - Educational Effectiveness

Chairperson: Anneke Timmermans, University of Groningen, Netherlands

University-based restructuring as organizational development: Actors' perceptions of effectiveness

Keywords: Curriculum change and development, Effectiveness and efficiency, Teacher professional development, Higher education

Presenting Author:Katharina Hellmann, University of Education Freiburg, Germany; **Co-Author:**Stefanie Vigerske, University of Education Freiburg, Germany

Teachers' professional development and universities' influence on creating competent teacher candidates are topics of continuing interest both in scientific research and policy. Based on this, site-specific concepts for teacher education are implemented regularly at universities, yet the evaluation and perpetuation of newly developed structures, teams, or curricular changes are usually pending. In this contribution, results of focus group discussions with the professorial and administrative level are presented that covered the topic of university-based organizational development, its impact on various levels of the organization, and university actors' criteria for an effective teacher education. Background to this is a recently started cooperation of two universities onsite that, so far, were only loosely connected. The possible persistence of the newly developed structures, the interviewed actors' perspectives of educational effectiveness and feasible criteria for determining "effectiveness" were at the centre of the discussions. Results of these focus group discussions will be presented, with a special focus on the partially divergent views of the professorial and administrative level on effectiveness. Implications for further restructuring onsite will be presented and the need for a continuous evaluation and quality management system will be outlined.

Cooperation and Collegiality and Readiness for Innovation in German Primary All-day Schools

Keywords: Culture, Social interaction, Teacher professional development, Quantitative methods

Presenting Author:Frederik Osadnik, Institute for School Development Research (IFS), TU Dortmund, Germany; **Co-Author:**Karin Lossen, Technical University of Dortmund, Germany; **Co-Author:**Wolfram Rollett, University of Education Freiburg, Germany

This paper examines the impact of teacher cooperation and collegiality on their readiness for innovation. By applying multilevel longitudinal analyses, we study the interrelations of these variables on both the individual level and the school (organizational) level. In order to do so, we use the national data set of the "Study on the Development of All-Day Schools (StEG, Holtappels et al., 2008)" which is representative for the development of all-day schools in Germany between 2005 and 2009. For our analyses we use the survey data of 859 teachers from 75 primary schools over 4 years. In the two-level cross-lagged-panel-model the overall results show that a more cooperative school climate and its improvement over time leads to a more favorable development concerning the readiness for innovation on both levels.

Organizing sustainable in-school professional development: lesson study as an organizational routine

Keywords: Program implementation and evaluation, Teacher professional development, Secondary education, Case studies

Presenting Author:Fenna Wolthuis, University of Groningen, Netherlands; **Co-Author:**Siebrich De Vries, University of Groningen, Netherlands; **Co-Author:**Klaas van Veen, University of Groningen, Netherlands

Lesson study (LS) is a school-based professional development program which originates in Japan where it has been an integrated part of the educational system for 130 years. LS has spread to other countries around the world interested in gaining similar advantages for teacher learning and student achievement. It is a promising program as it contains all of the features which are known to make teacher learning effective, namely long-term, coherent, collaborative, classroom and school-based professional development. Although research has shown that LS can be effective outside of Japan, many countries struggle to meaningfully sustain LS within schools. In this study we investigate if and how fifteen schools of which between 1 and 4 teachers participated in a cross-school professional learning community with the purpose of becoming lesson study facilitator at their own school, created lesson study as an organizational routine and which factors hinder or promote this process. Results show schools differ substantially in the extent to which they create and develop LS as an organizational routine. This was influenced by 1) facilitation and support from school leadership, 2) both teachers' initial motivation to participate in the LS-PLC and the motivation of their colleagues to participate in lesson study in the school, 3) the extent to which teachers perceived LS to be relevant for their work.

Session Fri 31, 11:00 - 12:30 4

31 August 2018 11:00 - 12:30

Auditorium

Symposium

How do schools change? –Applying innovative methods to investigate conditions of school improvement

Keywords: Case studies, Class and school composition, Developmental processes, Effectiveness and efficiency, Improvement, Leadership, Mixed-method research, Motivation, Quantitative methods

Interest group: SIG 23 - Educational Evaluation, Accountability and School Improvement

Chairperson: Markus Sauerwein, German Institute for International Educational Research (DIPF), Germany

Chairperson: Sebastian Wurster, Johannes Gutenberg University Mainz, Germany

Discussant: Tobias Feldhoff, Johannes Gutenberg University of Mainz, Germany

Schools are organizations in continuous movement and change. They must adapt to various external requirements, but are also encouraged to develop themselves on their own initiative. Effective schools in improving teaching and learning are able to develop a stimulating climate, implement tools and processes that enable the school actors to adapt current visions, goals and strategies. They reflect upon their own movement and changing processes. However, serious research gaps can be identified. School improvement research was mostly based on standardized questionnaires. Sound research to understand mechanisms of school improvement needs longitudinal designs and methods to investigate improvement processes while they happen. The symposium consists of three studies focusing on teachers within improvement processes. They investigate different conditions of successful improvement processes with innovative methods. The first paper focuses on teacher regulation activities. Therefore teachers filled in a daily online practice log at the end of every workday. The second paper gains a view insight the work of school improvement groups. Transcripts of the meetings from school development groups are analyzed by text mining and combined with students questionnaires. The third paper examines how teachers' engagement in professional learning activities could be enhanced. In this longitudinal study latent change score models are applied. Based on all studies promoting and hindering conditions for school improvement processes as well as methodological issues of school improvement research are discussed.

Regulation processes of teachers in secondary schools.

Presenting Author:Katharina Maag Merki, University of Zurich, Switzerland; **Co-Author:**Urs W. Grob, University of Zurich, Switzerland; **Co-Author:**Ariane Rickenbacher, University of Zurich, Switzerland; **Co-Author:**Beat Rechsteiner, University of Zurich, Switzerland; **Co-Author:**Andrea Wullschleger, University of Zurich, Switzerland

How do school development teams implement peer mentoring programs?

Presenting Author:Markus Sauerwein, German Institute for International Educational Research (DIPF), Germany; **Co-Author:**Julia Dohrman, German Institute for International Educational Research (DIPF), Germany

Fostering sustained teacher learning

Presenting Author:Arnoud Oude Groote Beverborg, Johannes Gutenberg University of Mainz, Germany; **Co-Author:**Peter Slegers, BMC, Netherlands; **Co-Author:**Nienke Moolenaar, Utrecht University, Netherlands; **Co-Author:**Klaas van Veen, University of Groningen, Netherlands

Session Fri 31, 12:45 - 13:45 1

31 August 2018 12:45 - 13:45

Auditorium

SIG 18 - 23 Keynote Session

SIG 18 - 23 Keynote Session "25 years of school leagues tables, accountability and choice: Lessons from England"

Keywords: Accountability and school inspections, Effectiveness and efficiency, Hierarchical linear modeling, Quantitative methods

Interest group:

Chairperson: Roel J. Bosker, Rijksuniversiteit Groningen, Dept of Education and GION, Netherlands

25 years of school leagues tables, accountability and choice: Lessons from England

Keywords: Accountability and school inspections, Effectiveness and efficiency, Hierarchical linear modeling, Quantitative methods

Presenting Author:George Leckie, University of Bristol, United Kingdom

Education systems around the world increasingly use school performance measures derived from value-added statistical models to estimate school and teacher effects on student test scores and to reward and penalise schools and teachers accordingly. England has a 25-year history in this regard, additionally publishing schools' performances in high profile 'league tables' to facilitate parental choice. In this talk, I will describe the evolution of school performance measures and

statistical models brought about by successive English governments. I will then discuss the statistical challenges England has encountered when attempting to estimate school effects, including: measuring their stability and consistency, quantifying and communicating their statistical uncertainty, accounting for student mobility and missing data, and monitoring their influence on narrowing student achievement gaps. I will conclude with thoughts and recommendations for other education systems embarking on data-driven school accountability systems.