

Attachment 8

Designing PISA as a Sustainable Database for Educational Policy and Research:

The PISA 2012 Context Questionnaire Framework

(EDU/PISA/GB(2010)23)

TABLE OF CONTENTS

DESIGNING PISA AS A SUSTAINABLE DATABASE FOR EDUCATIONAL POLICY AND RESEARCH THE PISA 2012 CONTEXT QUESTIONNAIRE FRAMEWORK.....	4
INTRODUCTION.....	4
1. General purpose and policy relevance of PISA.....	6
1.1 Comparative system monitoring.....	7
1.2 Understanding patterns of effectiveness in education systems.....	8
1.3 Building a sustainable database for policy relevant research.....	10
1.4 Focus on equity.....	11
2. The general knowledge base: Research in educational effectiveness.....	11
2.1 A rubric of educational outcomes and predictive factors.....	11
2.2 System level inputs and processes.....	14
2.3 School-level inputs and processes.....	17
2.4 Classroom level inputs and processes.....	19
2.5 Student-level inputs and processes.....	21
3. Learning conditions for mathematical literacy.....	23
3.1 Mathematical literacy: A challenge for instruction and assessment.....	23
3.2 Outcome measures related to mathematical literacy: Strategies, beliefs, and motivation.....	24
Strategies and metacognition.....	24
Motivation and intentions.....	25
Self-related beliefs and planned behaviour.....	26
3.3 Opportunity to learn and quality of instruction: Assessing the learning environment.....	27
Opportunity to learn – Coverage of content categories and problem types.....	28
Learning time.....	29
Quality of instruction.....	29
Teaching practices and students’ mathematical activities.....	30
3.4 School- and system-level support for teaching and learning of mathematical literacy.....	31
4. Specifying the questionnaire design for PISA 2012.....	32
4.1. Distributing content across PISA cycles: An overarching design for general, domain-specific, and thematic extension variables.....	32
(I) General variables (for all cycles).....	32
(II) Domain-specific trend variables (for major domain only, included every 9 years).....	32
(III) Thematic extension variables (extensions within individual cycles).....	33
(IV) System-level data, gained from INES or system-level questionnaire.....	33
4.2 Exploration of measures from PISA 2003.....	34
4.3 Rotation design for extended student questionnaires and computer-based delivery.....	36
Rotation.....	37
Computer-based delivery.....	37
5. Assuring validity evidence, explanatory power, and policy relevance of the PISA design.....	38
5.1 How to model the effects of context variables on student learning and performance.....	38
5.2 Cross-cultural validity.....	41
Item bias.....	41
Method bias.....	41

5.3	Extending the sample to study change at the school level (international option).....	42
5.4	Longitudinal extensions at the student level (recommended for future cycles).....	43
5.5	Final remark.....	44
References.....		45

Tables

Table 1.	Two-dimensional taxonomy of educational outcomes and predictive factors.....	14
Table 2.	A classification of PISA 2003 questionnaire.....	36

Figures

Figure 1.	The theory of planned behaviour.....	27
Figure 2.	An example of mediated moderation.....	40

**DESIGNING PISA AS A SUSTAINABLE DATABASE
FOR EDUCATIONAL POLICY AND RESEARCH**

THE PISA 2012 CONTEXT QUESTIONNAIRE FRAMEWORK¹

Introduction

This document provides an updated draft of the PISA 2012 context questionnaire framework following feedback from the PGB after the cancellation of the PGB meeting in Copenhagen and following further discussions in the Questionnaire Expert Group. It revises the draft that was prepared for the Copenhagen meeting [EDU/PISA/GB(2010)8].

The PGB is invited to:

- **FINALISE** the PISA 2012 context questionnaire framework.

INTRODUCTION TO THE FRAMEWORK

Throughout the first decade of the 21st century, the OECD's Programme for International Student Assessment (PISA) has been known for its outstanding contribution to policy discussions within OECD and partner countries and economies. Literacy and life skills of adolescents, assessed at the age of 15, capture the yield of years of learning both within and outside schools. Reading, science and mathematical literacy are important prerequisites for individual success in academic, professional, civic, social, and private contexts. When reported by systems or countries, the average level of those skills is a major indicator of human capital, which in turn impacts upon the prosperity and well-being of society as a whole. How these skills are distributed across the population, dependent on gender, ethnicity, socio-economic background, or geographic region, will help to understand diversity and equity within and between countries. The performance data delivered by PISA have been combined with information on student backgrounds to provide information on level, diversity and equity of life skills. Policymakers, researchers and lay audiences alike compare what their respective country has achieved to other countries, using PISA to set benchmarks. In order to elicit the necessary background information, students participating in any of the PISA cycles have been questioned in detail about their parents' education and occupation as well as their own immigration background, and gender.

Literacy and life skills are to a large extent products of family and school education. In order to describe these learning contexts and to help understand their impact on student performance, PISA has also asked students, school principals, and sometimes parents about their perspective on the learning context. This additional information allows PISA to answer questions such as: Do countries differ in terms of instructional culture, *i.e.* can we identify country-specific profiles of practices in teaching? Do different

¹ . Drafted by Eckhard Klieme, Eduardo Backhoff, Werner Blum, Jack Buckley, Ying-yi Hong, David Kaplan, Hank Levin, Jaap Scheerens, William Schmidt, A.J.R (Fons) van de Vijver, and Svenja Vieluf

schools within countries serve different student populations and, if so, are their policies and practices targeted towards their respective populations? In connection with student performance data, the question may be asked: Can performance differences between individual students be accounted for by differences in societal characteristics, families, school resources, instructional practices and communities? The answers, suggested by PISA, have been highly influential in debates on quality, efficacy and efficiency of education systems around the globe. Within one decade, PISA 2000, 2003, 2006 and 2009 have established one of the most important global databases that may be used to assess educational effectiveness. Policymakers may use the answers they obtain from PISA to guide their decision-making by re-defining goals, setting priorities for educational reform, re-allocating resources, promoting new pedagogical concepts and organisational strategies and eventually improving education. PISA increasingly guides change in educational institutions and systems in many countries. As it re-evaluates education systems every three years, PISA is a key to monitoring change in education.

Given that PISA is entering its fifth cycle of surveys, it is time to reconsider the overarching framework that guides the development of questionnaires and the way that this information is used in analysing and reporting data. Between 2000 and 2009, the PISA student questionnaires, for instance, have in total gathered data on 71 scales covering family resources and activities, the learning environment at school and classroom level as well as student beliefs and motivations. However, none of these scales has been administered in all four cycles, while 58 scales have been used just once. Thus, each of the PISA cycles has been more or less designed as an independent study on learning contexts that focuses on a specific major domain. Questionnaire content has largely been determined by contemporary priorities rather than long-term policy and research goals. Undoubtedly, this design principle was useful as long as the survey was still being developed and new domains or innovative measures continued to emerge. Certainly, the design and analytical framework of the survey will continue to adapt to changes in society and education, such as the increasing importance of information and communication technology (ICT) both as a learning tool and a domain of cross-curricular skills, or the growing need to combine multiple settings for student learning (schools, after-school programs, e-learning, or even home schooling). At the same time, however, PISA has matured. As a sound and stable basis for international comparative studies and trend information on education systems, PISA now needs a coherent overarching architecture for future context questionnaires, which balances stability with innovation, and domain-specific with general issues. The current document aims to develop such architecture as a major foundation for the design of the PISA 2012 context questionnaires.

The document is organised into five main sections:

- Section 1 elaborates on the general purpose and policy goals of the PISA study and explains why, and what kind of, context information is needed to serve these goals. As a tool for comparative system monitoring (1.1), as a study that helps to understand the effectiveness of educational policies and practices (1.2) and as a database for policy related research (1.3), PISA has to address various factors on the student/family, school and system level. Equity issues warrant special concern within educational policy (1.4).
- Section 2 establishes the general knowledge base that PISA can use to fulfil its functions. The section starts by establishing a rubric that helps to organise constructs and measures that operate at different levels (2.1). This taxonomy is based on research in educational effectiveness, which is subsequently outlined in some detail to justify the selection of input, process and outcome measures at the system, school, classroom, and student levels (2.2 – 2.5, respectively). General findings on educational productivity and effectiveness are also presented here.
- Section 3 deals with the study of learning environments in a way that is relevant to the assessment of mathematical literacy as the focal domain in PISA 2012 (3.1). The questionnaires

will be designed to add further outcome measures such as mathematical attitudes and strategies (3.2), indicators of instructional quality, including the possibility that students were exposed to the problem types used in PISA (*i.e.* Opportunity to learn) (3.3), as well as school- and system-level context variables (3.3).

- Grounded in this knowledge base and directed by the policy purpose of PISA, Section 4 delineates the PISA 2012 questionnaire design. Three steps appear requisite: Firstly, an overarching architecture for constructs and measures for future PISA cycles is established (4.1). Secondly, in order to put earlier work to appropriate use and to establish trend information, measures used in PISA 2003, which also focused on mathematics as the major domain, are revisited and measures of proven quality and relevance are identified (4.2). Finally, implementation issues are discussed that are new to PISA (4.3): Applying a rotated design to the student questionnaire will allow for more material to be used in the study. Some questions will be answered by all students, as in previous cycles, some by sub-samples of students, while computer-based delivery of the school questionnaire will improve the user-friendliness and adaptability of questionnaire administration.
- Section 5 examines principles of analysis and design that may help to keep PISA innovative and to ensure the validity of measures in the future, *i.e.* complex statistical modelling techniques to define and identify effects (5.1), assuring cross-cultural equivalence and validity of the measures (5.2), re-sampling schools that took part in PISA 2003 to capture change at the school level (5.3) and using follow-up tests or longitudinal extensions to address suitably the added value of schooling (5.4).

1. General purpose and policy relevance of PISA

PISA serves an array of parallel purposes. The views endorsed by different stakeholders in the participating countries may be broken down into the following broad areas:

- PISA is a monitoring structure that provides reliable comparative information on education systems, describing system structures as well as the functioning and the productivity (*i.e.* the gross outcome or “yield”) of education systems. PISA data cover student career paths up to secondary level, school characteristics, school governance, student performance and motivation as well as equity issues (such as performance by gender as well as socio-economic background).
- PISA is an international study contributing to our knowledge base on educational effectiveness. PISA observes patterns of relationships between inputs, processes and outcomes of education. Thus, it helps to understand how educational outcomes are produced. Firstly, PISA allows for a decomposition of variation of student performance by the individual, school and system. Moreover, each PISA cycle provides data about multiple factors – covering these three levels – which, according to previous research, are expected to impact performance in reading, mathematics, or science. In addition to describing these factors, PISA estimates their direct and indirect relationships to student performance and other outcomes. Large representative samples and sophisticated statistical models allow for the generalisation of findings both within and across countries.
- PISA provides a data source for the study of educational contexts in general (*e.g.* how family, school and out-of school education interact in the development of life skills) and the study of educational variables in economic and sociological contexts (*e.g.* the relationship between demographics, economic wealth, economic growth and human resources). The database will

become even more informative with the availability of trend data also for mathematics and will cover a dozen years as soon as PISA 2012 will be included.

Thus, PISA offers three types of policy-relevant “products”: (1) *Indicators* that monitor the functioning, productivity and equity of education systems. (2) *Knowledge* on factors that determine educational effectiveness. (3) A reliable, sustainable, comparative *database* that allows researchers world-wide to study basic, as well as policy-oriented, questions.

1.1 Comparative system monitoring

A central goal of PISA is to monitor education systems in terms of student performance (literacy, or more generally, life skills), as well as non-cognitive outcomes (such as student motivation and well-being), educational careers, context variables (such as the students’ cultural, ethnic and socio-economic backgrounds) and, finally, process characteristics at the school and system levels (including evaluation and accountability policies, student selection and allocation, parental involvement, staff cooperation and opportunities to learn). PISA also provides indicators that cover relationships between these factors, *e.g.* the so-called social gradient which measures the strength of the relationship between socio-economic status and performance, or the relationship between educational resources, and outcomes. The most important use of PISA data concerns the development of input, process, outcome, and relational indicators that can supply the OECD’s education indicators programme. These indicators, in turn, trigger public debate, shape educational policy and inform decision-making.

The policy relevance of this system monitoring enterprise is based on (a) defining and operationalising cognitive and non-cognitive outcome measures that inform the selection and prioritisation of educational goals within participating countries, (b) examining and reporting factors that may be subject to control by policy and professional practice (so-called malleable factors) and (c) providing international benchmarks that allow policymakers to ascertain what they may learn from other countries. The selection of indicators is generally guided by policy demands. Educational policymaking must deal with the functioning of the school system (*i.e.* operational characteristics such as resources allocated to schools), with productivity (such as the gross level of student outcomes) and, last but not least, with equity (*e.g.* how resources are distributed).

Reporting policy relevant indicators requires not only the assessment of performance data such as students’ mathematical literacy, but also data based on student, school and parent questionnaires covering a broad range of context, process and non-cognitive outcome measures. This document provides theory and policy arguments that guide the selection of constructs and the definition of questions and scales.

Prior to PISA, a number of quantitative and qualitative studies in comparative education have provided insights into the history, the functioning and, to some extent, the effectiveness of education systems. However, PISA is unique in combining the following features:

- It provides rigorous data and an integrated set of indicators for the monitoring of education systems.
- It has a clear focus on yield, in terms of student performance measured at the end of compulsory schooling. It also intends to report on malleable features of education systems and institutions that may initiate policy decisions, intervention, and improvement.
- It provides trend indicators, thus allowing for the description of changes in a country’s performance levels over time as well as the development of background variables, processes and non-cognitive outcomes and an examination of the relationships between these various constructs. The more PISA moves into repeated cycles of measurement, the more can be learned

from examining the stability and change of educational input, processes and outcomes, and their relationship over the years.

Multiple examples of indicators based on PISA context data can be found in recent editions of the OECD's Education at a Glance reports (OECD 2007a, 2008, 2009a), such as:

- Relationship between immigrant background and student performance (2007 A6);
- Profiles of top performing students, including their attitudes and motivation (2009 A4/A5);
- Patterns of student attitudes related to geographical region or language (2007 A5);
- Relationships between resources and outcomes in education (2007 and 2008 B7), especially with regard to class size (2008 D2);
- Outcomes of vocational versus general educational programs (2007 and 2008 C1);
- Use of evaluation and assessment in education systems (2008 D5);
- School quality from a parental perspective (2008 A5); and
- Relationships between student background and access to (or motivation to participate in) higher education (2007 A4/A7, 2008 A3/A7).

One of the most important challenges to the development of the contextual framework and the resultant questionnaires is to ensure that indicators can be compared across cycles, while at the same time allowing for new indicators to be introduced and established. After a decade of international student assessment in PISA, with context questionnaires mostly related to the respective major domain of performance, the time appears ripe to structure and order all the constructs and instruments that may be used either in single cycles or across cycles as a source of trend information. In the future, the policy relevance of PISA will most likely strongly depend on how well this challenge is met.

1.2 Understanding patterns of effectiveness in education systems

Indicators will direct public attention towards successful and less successful sectors of the education system, or to goals that have been met versus goals that still pose challenges. Thus, the main goal essentially is to guide priority setting and decision-making in educational policy. Besides *obtaining descriptions* of strengths and challenges with regard to student performance and the conditions of teaching and schooling in their respective countries, policymakers also want to understand *why* students achieve certain levels of performance. In serving an important explanatory goal, PISA context instruments are designed to help to answer this question. Therefore, the PISA questionnaires have to cover the most important inputs and processes of student learning at the individual, school and system levels. Statistical models, using these multi-level data, will help to understand the complex relationships of how these inputs and processes interact with student outcomes. If data on resources and costs are available, PISA may also help to understand efficiency, *i.e.* effectiveness in relation to investments.

During the first decade of PISA, each survey used variables that were specifically related to the respective main domain. For instance, reading experience, interest in reading, instruction in language classes etc. were important factors in 2000 and 2009. By contrast, PISA 2003 included attitudes towards mathematics and mathematical activities during mathematics lessons, while PISA 2006 explored various types of science instruction and science-related beliefs and interest. Disciplinary Climate in the Classroom and Teacher Support in the Classroom were measured with regard to test language instruction in 2000 and 2009, but

with regard to mathematics instruction in 2003. The operationalisation and analysis of data were domain-specific, but the constructs, the hypothesised links to student outcomes and the interpretation of data were informed by general theories of educational effectiveness.

Using these domain specific context data, a number of insights have been gained, with obvious importance for teaching practice and policy. Here are some examples from the first PISA Research Conference that took place in September 2009². While some examples are based on cross-national analyses others are based on PISA data from one country:

- Raising the general level of students' reading engagement represents a means of improving equity (Baye, Monseur, & Lafontaine, 2009).
- Resource conditions of the school as well as the extent to which the school encourages students to use their full potential are both significant predictors of mathematics performance in Korea (Kaplan, 2009a).
- Instructional time is strongly linked to mathematical performance in Swiss schools, except for schools with advanced requirements (Angelone, Moser, & Ramseier, 2009).
- In order to support students' development of scientific competency and their interest in science, a focused pattern of scientific study seems to be more successful than increasing the breadth of scientific activities (Kobarg, Seidel, Prenzel, McCrae, & Walker, 2009).
- The extent to which schools prepare for careers in science is strongly associated with their students' wish to pursue a science-related career (Lie & Kjaernsli, 2009).
- Japanese students are interested in inquiry-based learning, whereas science teaching at the upper secondary level does not cater to that interest. This finding could contribute to an understanding of why Japanese students in PISA show relatively low levels of positive attitudes toward science (Yasushi, 2009).

Although the analysis of PISA data can make important contributions to the knowledge base for educational policy and practice, there are limits that have to be taken into account. Most important is the fact that PISA is a yield study, assessing literacy and skills that have been accumulated over the lifespan, from early childhood through different levels of schooling until the age of 15 years. PISA does not ascertain how much learning has taken place in the secondary school at which a student is presently enrolled. Such an assessment would require that the student's performance level was ascertained at the time of entering his or her present school and compared with the same student's present performance. In so doing, one would obtain a measure of progress or "value-added" in performance associated with educational experiences in the particular school. However, we do not have measures of student performance prior to the ones measured in PISA at age 15 years. Teacher quality and its impact on student performance cannot be judged in PISA either – at least with the design that has been in place for over a decade. This is because a random sample of 15-year-olds is taken in each school rather than intact classes, thus precluding the collection of information regarding classroom level instructional strategies. Finally, in one out of five countries that participated in PISA 2006, the majority of students had only recently been allocated to the schools in question, prohibiting direct conclusions on school effects within these countries.

As Baker (2009) notes, the history of policy-making informed by international comparative studies has seen a number of short-cut conclusions, based on too simple hypotheses as to the causes of performance differences at the system level. Also, econometricians have studied a number of issues in educational

² . Abstracts and papers available at <http://www.pisaresconf09.org/>

productivity, but still most of this work is descriptive in nature and does not allow for causal inferences (see Hanushek & Woessmann, 2010).

It is exceedingly difficult to draw causal inferences, such as concluding that a particular educational policy or practice has a direct or indirect impact on student performance, based on an observational survey and assessment data of the kind collected in PISA (Gustafsson, 2007; Kaplan, forthcoming). If, for example, links between high student performance on the one hand, and school evaluation data being accessible to the public (as a school level policy) on the other hand were found, the design of the study would not warrant causal interpretation. This is because data on at least some potentially important factors, such as prior student performance, cannot be collected in PISA. As a consequence, such potentially important factors cannot be included in the analyses and therefore not controlled statistically. Thus, PISA cannot tell if this policy happens to be applied in high achieving schools, or whether the policy actually results in higher student performance. The fundamental problem is that, in the absence of random assignment to a treatment (a policy or practice), it remains generally unknown whether unobserved factors exist that are related to selection of the treatment and to the outcome of interest. The drawing of causal inferences thus relies on the researcher's willingness to make additional, often un-testable, assumptions. Some researchers may attempt to match control students to the treated students on some observed control variables and discard students for whom there are no suitable matches. The problem still remains, however, that selection may be driven by unobserved factors. In any event, it is essential that researchers both clearly state all assumptions made and, ideally, assess the sensitivity of their causal inference to the violations of these assumptions. (See Section 5 for a discussion of advanced analytical methods and designs intended to address some of these issues.)

Much of the value of the programme is based on a constant interplay between PISA as a monitoring survey and more rigorous kinds of effectiveness research done elsewhere. Correlational and other exploratory results from PISA may subsequently be tested in longitudinal, experimental, or intervention studies. Even so, factors that have been demonstrated to be relevant for educational effectiveness or efficiency in the research literature are prime candidates for continuous monitoring within PISA and for incorporation into the OECD system of educational indicators (see Section 2).

1.3 Building a sustainable database for policy relevant research

PISA's influence already reaches far beyond educational practice, policy and research. PISA data are increasingly used also by economists and social scientists to examine broader issues such as the impact of human capital on economic growth (Hanushek & Woessmann, 2009) or how to predict successful integration of migrant families (Stanat & Christensen, 2006). The PISA research network together with a competitive process for commissioning thematic reports based on PISA data is aimed at fostering the use of PISA data also in economics and sociology. This initiative is expected to lead to different kinds of analysis and broader policy recommendations.

Broadening the scope of PISA as a database for policy relevant research requires that general constructs such as student socio-economic status and immigration background are operationalised in a highly sophisticated way. It also implies that additional information and data may be required (*e.g.* on civic values and health). When conceptualising the content of PISA questionnaires, it should be clear that the database that PISA generates will meet demands – both in research and in policy decision-making – that go beyond the system of indicators that has been established thus far.

In the long term, one of the major benefits of the PISA database will be the availability of trend data. Educational production processes can hardly be understood from cross-sectional data, because it is practically impossible to include all relevant variables – including information on previous performance – within the design. However, changes in performance at the country level (*cf.* Gustafsson, 2007; Hanushek

& Woessmann, 2010) or even changes at the school (see Section 5.3) and the individual student level (Section 5.4) can be interpreted and explained substantively, once changes in input and process are available. This goal requires that PISA defines a general set of variables that will remain constant over several cycles in the future (see Section 4.1).

1.4 Focus on equity

Equity refers to equal or fair distribution of educational provisions, learning opportunities and outcomes among participants in education with different characteristics (OECD 2005b, p. 14). A rough criterion for equity may be low variation of cognitive and non-cognitive outcomes between and within schools. The gap between high achievers and low achievers should be as narrow as possible. In a more sophisticated approach to examine equity, specific groups are compared. These are often based on geography (*e.g.* regions or urban/rural distinctions), public/private school enrolment, socio-economic status, gender, ethnicity, immigrant status and the heritage language. Small outcome differences between such groups may indicate equity. Equity can further be assessed in terms of the distribution of access to schooling, learning resources and opportunities among these groups. Policies aiming at equity may either aim at an equal distribution of resources, or provide additional support for disadvantaged groups. Finally, a “meritocratic” notion of fairness accepts that diversity exists among students, but claims that differences in educational opportunities and outcome should not be grounded in social origin (Cleary, 1968). PISA provides data that help evaluate education systems with respect to these different criteria of equity.

In addition to assessing the degree of equity, PISA 2012 can also be used to examine characteristics of systems and schools, including educational policies, that are likely to be related to equity. At the system level, such analysis may include comparisons of systems that differ regarding tracking and the funding and management of schools. Policies aimed at promoting equity encompass equalising resources or providing targeted support for schools with a disadvantaged student population, as well as regulations on remedial instruction, support with language learning and other extra-curricular activities. Similarly, at the school level, ability grouping practices and practices to assist students with special learning needs appear relevant. Adaptive instruction as well as a supportive and well-structured learning environment can be related to equity at the classroom level. Moreover, the prevalence of such characteristics in more or less equitable systems can be compared.

PISA has included variables that are relevant for equity as essential variables since its first cycle was conducted in 2000. Thus, trends in equity can be analysed as well as associations with changes in educational policies and aggregated school processes. Information may thus be obtained regarding changes in equity, both among and within countries, covering more than a decade.

2. The general knowledge base: Research in educational effectiveness

2.1 A rubric of educational outcomes and predictive factors

Schools fill multiple functions within modern societies. Consider thus:

- Teaching and learning qualify young people for success in academic and vocational careers.
- Grades and certificates guide educational and professional careers and help allocate adolescents to career pathways that fit their profile of competencies and personal goals.
- The school as an institution with rules and norms – some explicit, some hidden – educates minds and hearts.

Altogether the school integrates young people from quite different backgrounds into society by allowing them to participate in academic work, in social interaction, in cultural activities, and in civic life.

PISA defines literacy as the capacity of students to use texts of various kinds, mathematical tools and science-based reasoning as they encounter, interpret and solve problems and make decisions in real-life situations. In modern, knowledge-rich societies, literacy is one of the most important goals and outcomes of schooling. PISA also attempts to measure broader cognitive skills such as strategies for learning and thinking and problem solving competency. Success in school – and in life – further depends on being committed, sharing values and beliefs, respecting and understanding others, being motivated to learn and to collaborate and being able to regulate one’s own learning behaviour. These constructs can be perceived as prerequisites of cognitive learning, but may also themselves be judged as goals of education, as the OECD project *Defining and Selecting Key Competencies* (DeSeCo) has elaborated (Rychen & Salganik, 2003). Therefore, PISA addresses non-cognitive outcomes like attitudes, beliefs, motivation and aspirations and learning-related behaviour, such as self-regulation, strategies and invested time. As one such non-cognitive outcome, truancy has received increased attention as an important (negative) indicator of student’s use of learning opportunities, which is also predictive of school drop-out and deviant behaviour (Kearney, 2008; Lee & Burkam, 2003). These non-cognitive outcomes are measured mainly within the student questionnaire but also in the school questionnaire.

A large section of the student questionnaire, the school questionnaire and the international options, however, is devoted to contextual factors that are linked to cognitive and/or non-cognitive outcomes. These are used to understand effectiveness and to define indicators. The factors can roughly be classified as being either input or processes. Input factors are mostly related to the individual’s social and personal background. Also, structural features like school size and funding are treated as inputs. Processes include learning and teaching as core processes with various variables designed to capture their quantity and quality. Moreover, professional activities by teachers and principals as well as school policies and practices are classified as process variables.

A particular challenge with international surveys is that contextual factors affecting student learning occur at four different levels, *i.e.* student/family, classroom, school and country. The basic production process goes from:

- Background and structural factors (inputs) through interactions, activities and policies (processes) to performance and non-cognitive outcomes: For instance, students’ mathematical performance supposedly depends on the profile of mathematics-related activities in the classroom, which again may depend on teacher competency, class size, availability of technology and other resources.
- Higher-order levels (system, school) to classroom to student level processes and outcomes: The issue of whether the locus of control is at the school or at some higher administrative level will partially determine instructional leadership and teacher collaboration, which shapes teaching strategies, and finally students’ learning activities.

The basic structure of the Input-Process-Outcome model was developed in the 1960s for the International Association for the Evaluation of Educational Achievement (IEA) (Purves, 1987). A recent version of this model, as shown in Table 1Error: Reference source not found, accommodates these dimensions. The first column displays four levels: Students, classrooms, schools and countries. The three production phases are then given in the remaining columns, *i.e.* inputs, processes and outcomes, respectively. Each cell contains several example constructs. The rubric is quite comprehensive: It includes most of the factors that have previously been studied in PISA and practically all constructs are covered in this framework. Although it is a selection from the broad set of constructs discussed in the educational effectiveness literature, the table is

still too comprehensive from a research design perspective. The number of constructs that will finally be able to be covered depends on the rotation design that will be applied (see Section 4.4).

Some input factors are fairly stable and difficult to change while others can be shaped by school development activities or policy decisions. Processes are usually more malleable, at least indirectly (*e.g.* by teacher education and professional development) and outcomes reflect the effects of the inputs and processes. Note, however, that the discrimination between the three strands of variables is by no means clear-cut: Outcomes from one educational setting become input for the next, while some process aspects (*e.g.* learning strategies) may well be treated as input or outcome, depending on a given theoretical perspective, research design, or practical considerations.

When turned into a dynamic model of school effectiveness (see Creemers & Kyriakides, 2008), outcomes become inputs for further development. Mathematics anxiety, for example, can be an outcome of schooling as well as an input, impacting, for instance, upon students' homework activities. Moreover, inputs may have reciprocal effects upon each other. For example, a school's socio-economic composition in many education systems is correlated with funding, parental involvement or even teacher quality. This, in turn, allows for other (better) teaching-learning environments to be offered, which attract students (or, rather, parents) from higher socio-economic backgrounds, so that finally social stratification, resources and process quality are mixed and are difficult to disentangle.

In order to understand these relationships and effects correctly and to build sophisticated and adequately specified analytical models, we have to refer to knowledge accumulated in educational research. Two areas of research are especially important: Research on educational effectiveness in general (discussed below) and domain-specific research on learning and teaching (see Section 3).

Table 1 describes a constellation of factors at different levels that constitute an education system. Their interactions are complex and not fully understood and each level is important in shaping educational policy and probable education outcomes.

The main goal of educational effectiveness research is to identify “factors in teaching, curriculum, and learning environment at different levels such as the classroom, the school, and the above-school levels [that] can directly or indirectly explain the differences in the outcomes of students, taking into account background characteristics, such as ability, socio-economic status, and prior attainment” (Creemers & Kyriakides, 2008, p. 12) However, practices may neither be equally effective for all students within a school nor for all education systems, local contexts and schools. Moreover, depending on the kind of outcomes emphasised, different conclusions may be drawn (Kyriakides & Tsangaridou, 2004). Hence, modern research into educational effectiveness also takes interactions with input into account and examines differential effectiveness and adaptive practices. A considerable amount of research has been carried out in this field (*e.g.* Creemers & Kyriakides, 2008; Scheerens, 2000; Teddlie & Reynolds, 2000). The following sections will summarise results for each of the cells of the taxonomy introduced in Table 1³. Also, ways to operationalise major constructs will be discussed and design considerations will be discussed at the end of each Section.

³ . In order to connect the present framework to earlier conceptual work on PISA, sections 2.2 – 2.5 draw strongly on the PISA 2009 Questionnaire Framework, which was authored by Hank Levin and based on extensive work done by Jaap Scheerens. The authors gratefully acknowledge their important work as well as contributions made by other experts in previous PISA cycles.

Two-dimensional taxonomy of educational outcomes and predictive factors⁰

	Input	Processes	Outcomes
Students	Gender, grade level, socio-economic status Educational career, grades Immigration background Family environment and support ICT experience, attitudes, skills Openness, problem solving styles	Attendance/truancy Outside-class activities - e.g. participation in after school programs Motivation, engagement Learning and thinking strategies, test taking strategies Learning time (including homework and private tuition)	Mathematical performance Mathematics-related attitudes, beliefs and motivation General school-related attitudes and behaviour, e.g. commitment, truancy Learning motivation, educational aspirations
Classrooms	Class size, socio-economic background and ethnic composition Teacher education /training, expertise	Quality of instruction: structure, support, challenge Opportunity to learn: implemented curriculum, assigned tasks, mathematics-related activities Instructional time, grouping, assessment and feedback	Aggregated student variables
Schools	Socio-economic background and ethnic composition Affluence of the community School funding, public vs. private School size Parental involvement	Achievement orientation, shared norms, leadership, teacher morale and cooperation, professional development Admission and recruitment policies, tracking, course offerings/school curriculum, evaluation Teacher-student relations, supportive environment	Aggregated student variables Promotion/retention and graduation rates Attendance
Countries (Systems)	Economic wealth, social (in)equality Diversity policies	School funding, tracking and allocation, policies for professional teacher development, support for special needs and language minority students, hiring and certification policies Accountability and evaluation policies, locus of decision-making	Aggregated student variables Average graduation level

2.2 System level inputs and processes

In most countries, decisions about educational policy and the allocation of resources are the responsibility of a national or federal entity. In more centralised systems, this may be a ministry with ultimate authority, while in other countries the functions may also be performed by federal state authorities or regional agencies at the district level. Although system level factors are organised in terms of input, processes and outcomes in this taxonomy, their role in analytical models most probably will be the role of moderating variables; *i.e.* system level factors have an impact on how input, processes and outcomes are related at the lower level.

Basic inputs at the system level concern the material resources and their distribution within a country. Studies have found that both within and across nations the level of per-capita income is a strong predictor of student performance (Baker, Goesling, & LeTendre, 2002). Affluence provides resources to facilitate educational performance (Baumert, Carstensen & Siegle, 2005). At the same time, educational outcomes may also influence a country's economic well-being. Measures of societal inequality are linked to educational inequality in terms of family resources and learning conditions. Moreover, some countries provide for similar funding per student, while others specifically support students from rural, impoverished, or immigrant backgrounds, or those with learning or physical difficulties. A country's specific pattern of educational investment may have implications for both the level of performance as well as equity in outcomes and access to career pathways. Another important aspect that is closely related to societal inequality concerns policies on immigration and educational diversity (Stanat & Christensen, 2006). In a large number of countries immigrants perform at lower levels in some of the major subjects, but in other systems no differences are found or immigrants perform at a higher level than average. Smaller performance gaps exist in countries that actively recruit highly educated immigrants. By contrast, a large proportion of immigrants with low socio-economic status may pose a special challenge to education systems. Research has further shown that language support policies successfully reduce disadvantages of students with a heritage language that is different from the language of instruction.

To describe school systems it is further important to examine policies regarding ability grouping and the responsibility for school management. Some countries allocate students to schools with an academic focus or a vocational focus, or a professional focus, or a technical focus, according to their academic performance and/or their preferences. In such tracked systems, not all students are provided with the same learning opportunities. Therefore these practices have been subject to a recurring debate on whether educational separation denies academic opportunities to those students whose intellectual development is delayed as well as to those students of lower socio-economic status (Levin, 1978). Comprehensive systems, where all students are required to undertake at least a common core of academic work, may be more efficient in terms of equity (OECD, 2007b). However, even within schools academic classes may be heterogeneous in enrolments or streamed by ability. Besides their degree of tracking, countries also differ regarding the age at which students are allocated to different school forms as well as the possibility of changing tracks. Last but not least, school entrance and promotion policies will have an impact on PISA results; because – other things being equal – student performance will increase if students arrive at higher grades earlier (Gustafsson, 2007).

Previous PISA cycles have shown that public funding and management is most common, but in a majority of participating countries, a certain number of schools exist that are mainly funded and managed by non-governmental organisations like churches or businesses. Schools may further be privately managed but predominantly publicly funded; *i.e.* so called “government-dependent private schools”. Although both kinds of schools exist within a government regulatory framework, private schools usually have considerably more freedom than public schools to make educational decisions. Thus they represent one form of decentralisation of decision-making. Choice among public and private schools is considered as a mechanism to create incentives for schools to compete for students in order to improve the overall performance of the education system (Belfield & Levin, 2002). However, previous PISA results suggest that the higher average performance of privately managed and government-dependent privately managed schools is mainly due to the advantaged socio-economic backgrounds of their students and corresponding composition effects (OECD, 2007b). Hence, similar to the degree of tracking, the percentage of private schools and their characteristics may also affect equity.

In recent years, many countries have been addressing the issue of how their schools can be more responsive to the needs of specific students, groups of students and communities. When decision-making is vested in a central authority such as the national government, it is often argued that teacher hiring and training and the curriculum are too rigid to accommodate the needs of schools with different student inputs.

For this reason, many nations have established a variety of reforms to decentralise their schools, placing educational decisions closer to the students and communities being served, in the hope that this will improve school performance (Hannaway & Carnoy, 1993). However, studies linking this aspect to educational outcomes have not provided support for a strong association (*e.g.* Schmidt & McKnight, 1998). Some secondary analyses of international student assessments suggest that school autonomy is beneficial if combined with top-down evaluation or centralised exams (Woessmann, 2006).

Especially in decentralised systems it may be important to establish school and student assessments that will ensure that all students meet important educational standards and are equally prepared for their future educational careers (Carnoy, Elmore, & Siskin, 2003). This idea is endorsed by several countries, but their approaches to evaluation and accountability still differ significantly. Some countries sponsor periodic systems of testing to gain standardised information on student and school performance at national, federal, or regional levels. Others require such evaluations, but leave it to the regional and local authorities to carry out testing and assessment, while a further group of countries lack systematic evaluations. Furthermore, accountability also extends to the consequences of evaluation results (Carnoy, Elmore, & Siskin, 2003; Koretz, 2008). In some countries, tests and surveys only serve to inform the general public and educational entities. But test results can also be used to create incentives for higher performance, or to identify weaknesses and provide assistance to improve schools. Different systems may have different consequences for providing incentives to learn, uniformity of what is taught and educational results.

In addition to these rather descriptive variables at the system level, a number of more specific policies and practices have been proposed by the dynamic model of educational effectiveness (Creemers & Kyriakides, 2008). These aim at supporting schools realising a positive learning environment, as well as high quantity and quality of instruction and to offer students a large variety of learning opportunities. Few studies have integrated these system level processes. PISA 2012 could provide an opportunity to examine their prevalence and association with performance across a large sample of countries. Importantly, both the existence of these policies and the nature of their implementation might be analysed.

The learning opportunities offered are to a large extent determined by national curricula. The IEA's Trends in Mathematics and Science Study (TIMSS) identified substantive variation in the intended and implemented curricula across countries, especially with regard to time invested, coherence and focus, which in part explains the observed differences in student performance at least for curriculum-oriented tests like TIMSS (Schmidt *et al.*, 2001; Schmidt & Houang, 2007; see also Section 3.3 below). Yet, in addition to formal learning, policymakers can also support and encourage the provision of extra-curricular learning opportunities. They may, for example, launch mathematics competitions or provide schools with resources and support for establishing after school programs. Countries may also pursue different policies with regard to commercial tuition and other kinds of "shadow education" (Baker & LeTendre, 2005).

The quality of instruction can be controlled at the system level by defining standards for teaching and making sure that these standards are met. This implies the formulation of explicit expectations based on the latest research findings. Moreover, it presupposes that evaluations do not exclusively focus on measuring output but also examine teaching and instruction. By establishing respective policies, and organising and providing resources for in-service training, policymakers can further contribute to the professionalisation of school staff. Professional development is defined as "activities that develop an individual's skills, knowledge, expertise, and other characteristics as a teacher" (OECD, 2009b, p. 49). In many countries, it is compulsory for teachers to spend a certain number of days on professional development. However, education systems vary significantly regarding the general level of participation and the pattern of attendance in different types of professional development. Participation may be encouraged by financial support, salary supplements, and scheduled time. However, the OECD's Teaching and Learning International Survey (TALIS) results suggest that the effects of these measures are limited (OECD, 2009b).

Finally, cultural norms and values also influence the behaviour of different stakeholders. Students' and parents' engagement in schools and learning partly depends on the general appreciation of education and related values. The overall status of teachers is important because it affects the perceived attractiveness of the profession for people considering taking up teaching as a career. Community involvement in schooling can provide greater sensitivity through local adaptation to the needs of local populations.

Design considerations: Some of the system level variables, such as economic wealth, educational investments, equity of their distribution, composition of migrant populations, tracking and school entrance rules, may be captured from existing databases, *e.g.* from the OECD system of indicators on economy, welfare, and education (see *e.g.* http://www.oecd.org/document/7/0,2340,en_21571361_31938349_36043527_1_1_1_1,00.html). Others could be covered by a system level questionnaire to be answered by senior administration officers, such as policies for teacher recruitment and professional development, evaluation and accountability, school funding, centralisation/ decentralisation, standards and extra-curricular activities, support for special needs and minority students. Some indicators may be derived from aggregated PISA data, like the index for school autonomy in decision-making that is derived from the school questionnaire. Student, teacher, and parent data may be aggregated at the country level to provide measures for cultural contexts, such as appreciation of education relative to other personal and social assets.

2.3 School-level inputs and processes

The organisation of schools is complex and it varies considerably, not only from one country to the next, but also among educational sub-divisions within a country such as federal states or provinces, regions, school districts and individual schools. In discussing inputs at the school level, first of all the type of student input is likely to affect educational outcomes. Schools recruiting students of higher socio-economic status often tend to create general school environments of high aspirations and support for academic learning. Moreover, such schools may attract better teachers. This provides advantages even for students of lower socio-economic status who are enrolled in these schools. The student input is likely to depend on the general affluence of the surrounding community which has additional effects on student learning. Wealthier neighbourhoods tend to be less preoccupied with the possible effects of crime and violence and they have more resources to support schools in informal ways. Reciprocally, good schools reflect and may attract engaged parents and thus also influence the community. A somewhat different dimension is the proportion of students with a heritage language different from the language of instruction. Linguistic diversity is, on the one hand, a resource for schools. On the other hand, it detracts from resources since specific support systems and adaptive teaching skills may be required to deal with comprehension difficulties and likely cultural differences. In many countries, the socio-economic context varies greatly across urban, suburban, and rural communities.

When schools encourage and provide for parental involvement, parents can become more effective in supporting both school programs and the educational progress of their children. Furthermore, parents who participate in school activities are more likely to volunteer their efforts in assisting the school, therefore increasing available resources. Parents who know what schools expect by being familiar with the school program and teacher expectations are better able to assist their own children in learning. Finally, the act of attracting parents to participate actively in educational and school endeavours can serve to form social networks where parents get to know and help each other. Arguably, such social networks raise overall performance through the accumulation of "social capital" (Coleman, 1988).

School size is also thought to be linked to performance. With larger enrolment numbers, schools can offer their students a greater variety of teachers, courses, electives and extra-curricular activities. This enables students to select those courses that are most meaningful to them educationally and for which they are most motivated. Larger schools, however, also tend to be more impersonal and students may experience less

individual support. Research has thus shown that smaller secondary schools demonstrate a greater and more equitable distribution of student engagement and performance among 15-year-olds (Coleman, 1988). Indeed, the question about optimal school sizes remains unanswered. In fact, some research suggests that school size affects different groups of students (*e.g.* those based on socio-economic status) in different ways (Lee & Smith, 1997).

In addition to these inputs, a number of processes directly or indirectly predict outcomes at the school level. Arguably, the most important malleable characteristic of effective schools is the school climate (or the environmental quality of the school). The school climate encompasses not only norms and values but also the quality of relationships and the general atmosphere. An academic focus – a general consensus about the mission of the school and the value of education – facilitates learning. In addition, an orderly learning atmosphere maximises the use of learning time. By contrast, disrespectfulness and an unruly environment are counterproductive for both teachers and students alike and distract from the actual mission of the school. An orderly environment might, in turn, be fostered by coherent, reliable rules, (*e.g.* dealing with student misconduct like absenteeism). Effective schools are able to react to the specific learning needs of students and the needs of the teaching staff. They provide remedial and enrichment classes for students with learning disabilities, for highly gifted students and for students with a heritage language different from the language of instruction. In addition, effective schools offer student counselling and homework assistance. More broadly, the variety and quality of extra-curricular programs can also facilitate student learning and may be especially relevant for non-cognitive outcomes and students' future career paths.

Three school organisational aspects are relevant for improving teaching and the learning environment: Collaboration among staff, professionalisation, and school leadership. Collaboration among staff provides practical and emotional support for teachers and contributes to their professionalisation. It encompasses different techniques that aim at coordinating practice, making the curriculum more coherent, and providing mutual support and feedback. In many countries it is relatively common for teachers to exchange instructional material or to discuss the learning difficulties of individual students. More sophisticated forms of cooperation include collective learning activities such as observation visits, providing feedback, engaging in professional learning activities and joint activities across classes and age groups (OECD, 2009b). The latter practices in particular help to transform schools into learning organisations as well as providing constructive feedback for teachers and support for professional development activities specifically addressing teachers' needs. Principals or school heads need to deal with administrative tasks, such as legal and budgetary issues, facility management, and public relations. However, the core of their work is to assure a high quality of instruction and learning within their schools. During the last decades, the concept of instructional leadership, which puts special emphasis on this aspect, has received considerable attention in the research literature (Blase and Blase, 1998; Hallinger and Heck, 1996; Heck and Hallinger, 2005; Krüger, Witziers and Slegers, 2007; Leithwood and Riehl, 2005; Witziers, Bosker and Krüger, 2003). However, empirical results regarding the relationships of school leadership with student performance are inconsistent, possibly because leadership can have a rather indirect effect that is mediated by other school and classroom level processes.

The quantity of instruction can further be addressed by policies and regulations on lesson schedules, timetables, management of teaching time, student attendance and parent involvement. Guidelines on teaching, grouping procedures, and teacher behaviour can help raise the quality of teaching. Opportunity to learn is strongly influenced by the curricula and the selection of instructional materials. However, the formal descriptions of educational operations in schools are often misleading because implementation differs so widely (Fullan, 1992; Fullan & Stiegelbauer, 1991). Accordingly, the PISA questionnaires attempt to obtain information on operations directly from the participants, and especially the school principal.

It is insufficient to simply introduce the aforementioned policies and practices as their actual effects should also be monitored. School evaluation can be used to identify strengths and weaknesses in the school's operations, monitor practices and assess their effects on outcomes. TALIS has shown that evaluation practices vary widely across countries, not only with regard to the frequency but also concerning the sources of information used, the focus of evaluations and the consequences enacted (OECD, 2009b). So far, little research examines the effects of school evaluations on outcomes (Creemers & Kyriakides, 2008).

A somewhat different characteristic of schools that impacts decision-making is the degree of centralisation (versus autonomy) of governance. In some societies, the recruitment of teachers, development of curricula, and assessment of students is undertaken centrally. In other societies, the school is merely expected to follow directions set out for routine school operations. Still further, at the opposite end of the spectrum, the school is largely autonomous and is expected to choose its own teachers, to influence teacher compensation and to choose (and undertake) its own academic assessment of student performance. In recent years, school autonomy has been in focus as a way of overcoming bureaucratic rigidity and potentially impacting on learning outcomes (Bottani & Favre, 2001; Chubb & Moe, 1990). However, results are not yet conclusive and decentralisation is still subject to debate.

Design considerations: Thus far, most school characteristics have been addressed by asking the principal to comment on school resources (or the lack thereof), school curriculum (*i.e.* timetables, tracks, remedial and enrichment classes, extra-curricular activities), school climate (*i.e.* aspiration, teacher and student morale, parent involvement, behavioural problems), and professional activities (*i.e.* teacher collaboration, shared norms, leadership, evaluation procedures). Adding data from a teacher questionnaire, aggregated at the school level, would allow for a broader and probably more valid perspective on these issues. This approach is possible in those countries that link PISA 2012 to the next cycle of TALIS (OECD, 2009b). In addition to information from the school questionnaire on these issues, student data, including information from the optional ICT Familiarity questionnaire, and parent data can be aggregated as indicators of student composition, social capital, school resources and school climate (*e.g.* availability of ICT, teacher-student relations, parental attitudes towards and involvement in the school) while the optional Educational Career questionnaire addresses career counselling at school.

In PISA 2012, the school questionnaire can also be used to gather outcome data beyond the performance and conduct of those students who are directly tested. For instance, principals may provide data on student truancy, promotion and graduation rates. Such "objective" behavioural outcomes may be better comparable at least within an education system than most attitudinal or self-report measures. Also, they directly represent school-level outcomes while student performance is largely predicted by classroom- and teacher-level variables.

2.4 Classroom level inputs and processes

Most of a student's educational experiences at school occur in the classroom. Here, the student is exposed to subject content, curriculum materials, instructional strategies and to the specific composition of, and climate within, the class. As previous research has shown, proximal variables – among them classroom characteristics and practices – are more closely associated with student performance than more distal factors, such as school-level and system-level conditions (*e.g.* Wang, Haertel, & Walberg, 1993).

The most important inputs at the classroom level are the classroom context and the teacher (Wayne & Youngs, 2003). The former encompasses the characteristics of fellow students, instructional groupings, and class size. The concern about the socio-economic and language background of students at the classroom level is similar to that at the school level. In many cases, this composition varies not only between schools, but also within schools, particularly where grouping into different classes occurs according to students' performance levels. Although the rationale for such grouping is to enable teachers to improve instruction

by tailoring it specifically to student ability, there is considerable evidence that the educational impact of such grouping may widen the gap in performance between such groups (Gamoran, 1992; Oakes, 2005). Ability grouping within classes that is more flexible and occurs for certain tasks or certain periods of time only or heterogeneous ability grouping may, however, be useful to implement adaptive and cooperative classroom teaching practices (Slavin, 1990).

A specific aspect of classroom practice, which is arguably quite important in the context of assessments like PISA, is the everyday practice of student evaluation, assessment, and grading. Do students work on standardised tests regularly? Do they receive teacher feedback in terms of grades or written reports or oral feedback? Do schools, teachers, students and/or parents receive information about student performance as compared to general standards? How do they use this information? Do students practise test-taking? How do these practices evolve and change over time in different countries? As has been shown by research on formative and summative assessment practices (Brookhart, 2009), as well as in studies on conditions and effects of high stakes testing (Borko *et al.*, 2007; Koretz, 2008), student results in standardised assessments like PISA may interact with assessment-related policies and practices.

Class size is generally believed to have an (negative) effect on student performance. Smaller classes enable more opportunities to participate in class and facilitate the use of teaching practices that are targeted at individual (groups of) students. Moreover, fewer students mean that fewer assignments need to be evaluated leaving teachers more time for providing feedback and designing richer activities and assessments. And yet, international research suggests that class size reduction needs to be substantial before it has any effect, and it is less relevant in secondary schools compared to primary schools (Gustafsson, 2007). Moreover, effects of class size seem to be culture specific: Comparatively large classes are found in many Asian countries which nevertheless realise a high level of average student performance.

The teacher is characterised by her or his education and training, teaching experience and expertise. Advanced academic degrees, a major in the subject being taught, and professional experience have been described as desired qualifications and in certain models as indicators of teacher quality. However, results regarding their association with student performance are inconsistent (*e.g.* Hanushek & Rivkin, 2007; Libman, 2009; Mullis & Martin, 2007; Zuzovsky, 2009). Stronger effects have been found for teacher expertise (*e.g.* Baumert *et al.*, 2009; Hill, Rowan, & Ball, 2005). This has been defined as the interplay of teachers' general pedagogical knowledge, content knowledge, pedagogical content knowledge, beliefs, motivation and capacity to self-regulate (*e.g.* Bromme, 1997; Brunner *et al.*, 2006; Shulman, 1987). In addition to the inputs described above, a number of processes at the classroom level have been found to be relevant for educational effectiveness, particularly classroom climate and instructional quality. Research has shown that student learning is generally supported by a positive and respectful atmosphere that is relatively free of disruption and focused on student performance (Creemers & Kyriakides, 2008; Harris & Chrispeels, 2006; Hopkins, 2005; Scheerens and Bosker, 1997). The major facets of a positive classroom climate are: Supportive teacher-student interactions, good student-student relationships, achievement orientation, and an orderly learning atmosphere with clear disciplinary rules. Instructional quality, however, is a more complex aspect. Existing evidence suggests there is no single best way of teaching. Well-structured lessons with close monitoring, adequate pacing and classroom management, clarity of presentation and informative and encouraging feedback (*i.e.* the key aspects of "direct instruction") are linked positively to student performance. These components help create an orderly classroom environment and maximise effective learning time. Yet student motivation and non-cognitive outcomes benefit from additional characteristics of instructional quality, such as a classroom climate and teacher-student relations which support student autonomy, competency and social relatedness. Furthermore, in order to foster conceptual understanding, instruction has to use challenging content (Brown 1994). Also, different student sub-populations may benefit from different instructional practices. Thus, teachers have to orchestrate learning activities in a way that serves the needs of their specific class. Klieme, Pauli, and Reusser (2009) condensed this knowledge into a framework of three "basic dimensions of instructional quality": (a) clear,

well-structured classroom management, (b) supportive, student-oriented classroom climate, and (c) cognitive activation with challenging content. Several independent studies of secondary school mathematics education have now confirmed this triarchic structure of classroom quality and given some support for the cognitive and motivational impact that was hypothesised (*TIMSS-Video*: Klieme, Schümer, & Knoll, 2001; *COACTIV*: Baumert *et al.*, 2009; *Pythagoras*: Lipowsky *et al.*, 2009). Klieme and Rakoczy (2003, see also Kunter *et al.*, 2008) identified similar structures within national extensions to PISA. However, instructional quality, especially cognitive activation, is to a large extent domain specific. Aspects specifically related to mathematics learning will be discussed in Section 3.

Design considerations: With the exception of a few countries that extend their samples to include grade-based components the PISA samples do not include a classroom level as 15-year-old students are randomly sampled from all classes in a school. This sampling design presents challenges to the examination and classroom and teacher effects (Opdenakker & van Damme, 2000; Van Landeghem, De Fraine, & Van Damme, 2005) and ultimately the examination of well-specified models of educational effectiveness. In addition, it reinforces the yield nature of the PISA assessment. Still, PISA can report on classroom inputs and processes in a descriptive way, taking up variables that have elsewhere been shown to relate to student performance and assessing them by means of student, school, and (where possible) teacher questionnaires. For example, principals have responded to questions concerning teacher background and predominant pedagogical orientations while students have been questioned about classroom context and practices. The PISA scales Disciplinary Climate in the Mathematics Classroom and Teacher Support in the Mathematics Classroom are indicative for the structure and support dimension in the triarchic model of instructional quality, respectively. PISA 2006 added several indicators for advanced, challenging practices in science education, and PISA 2012 will do so for mathematics (see Section 3.3 below). Undoubtedly, details, for example, concerning teachers' constructivist vs. direct transmission beliefs and professional development activities would further add to this line of inquiry. For PISA 2012, this is likely to be achieved by some countries linking PISA 2012 schools to the sample for the TALIS teacher survey in 2013. In this way, such information, aggregated at the school level, would be available for secondary analyses of PISA data.

2.5 Student-level inputs and processes

By 15 years of age, it is not only the accumulated effects of schooling that contribute to an individual's academic performance, but also the experiences encountered at home. More educated parents are able to provide a richer set of learning opportunities. They are also able to provide more access to written materials for reading, travel, and other resources that engage their child's curiosity. Research has shown that parents holding high expectations for students' academic performance and showing interest in students' school work is linked to the educational success of their children, as are parents' participation in school conferences and involvement in homework (Alexander *et al.* 2007; Christenson 2004; Hoover-Dempsey & Sandler, 1997; Ma, 1999; Sui-Chu & Willms 1996; Wang, Haertel & Walberg, 1993). One of the main purposes of gathering data on the family background is to take into account these influences on learning when estimating school effects. In addition to the student questionnaire, the parent questionnaire as an international option collects information on several relevant variables such as career aspirations, parental involvement in school, and discussion of school related matters at home. Collectively, these instruments allow for the triangulation of parental support and academic expectations from varied perspectives. The ICT questionnaire provides data on computer availability and computer use – including at home –, related skills and attitudes, which may especially help to explain the results of computer-based tests of cognitive skills.

In addition to socio-economic and family background, the linguistic background of students appears relevant. Growing up bilingually may foster the development of language awareness and facilitate the learning of additional foreign languages (Hesse, Göbel, & Hartig, 2008). However, the possibly poorer competency in the language of instruction during the first years of schooling can also negatively affect

learning in other subjects and in the long run be disadvantageous for students' educational pathways (Schmid, 2001; Stanat & Christensen, 2006). On the one hand, immigrants and ethnic minorities may have to adjust to an unfamiliar cultural context at school (Berry, 1980; Berry, 1990; Hovey & King, 1996; Liebkind, 1996) and they may face discrimination (Amiot & Bourhis, 2005; Perreault & Bourhis, 1999). On the other hand, quite often these groups hold stronger aspirations and higher motivation than peers from ethnic majority families (Kao & Tienda, 1998; Krahn & Taylor, 2005; Stanat, 2006; Stanat & Christensen, 2006). PISA strives to identify the various inputs, processes and outcomes of migrant education in order to disentangle migration effects from socio-economic and language effects. Also, it seeks to separate individual level from compositional or contextual effects. PISA 2012 strives to extend the explanatory power of the study by including measures of acculturation, and perceived cultural proximity, as well as in-depth information on language background within the optional Educational Career student questionnaire. The optional parent questionnaire provides additional information on migration background while the optional educational career questionnaire provides additional information on linguistic background and upbringing.

General student-level processes mainly refer to the learning time. While formal learning opportunities are usually compulsory, students may still limit their use, *e.g.* by skipping classes, by arriving late or by being inattentive during lessons. Absenteeism and time spent on task significantly influence student performance and they are related with school drop-out rates, delinquency, and drug abuse (*e.g.* Baker, Sigmon, & Nugent, 2001; Lee & Burkam, 2003; McCluskey, Bynum, & Patchin, 2004; Wilmers *et al.*, 2002). Moreover, these behaviours are also relevant as an aggregate indicator of the school-level learning environment (see above). By contrast, students may extend their learning time and the variety of learning opportunities by individual study or by participating in extracurricular activities like academic clubs, competitions, volunteering and debating. Students with a disadvantaged social background seem to especially benefit from participation in such programs (*e.g.* McComb & Scott-Little, 2003). In some countries, taking private tuition and commercial courses, like Juku in Japan, is a common and important, though "shadowed" part of the education system (Baker & LeTendre, 2005).

Educational aspirations are a significant predictor of the students' future educational attainment (*e.g.* Thiessen, 2007), and they vary between different ethnic groups regarding their quality and stability (*e.g.* Kao & Tienda, 1998; Mau & Heim Bikos, 2000). Student motivation and attitudes towards learning are important student inputs, which may also be interpreted as non-cognitive outcomes. Just like students' learning styles these are to a large extent subject specific and will therefore be discussed in Section 3. General, domain-independent student characteristics, in contrast, can be used to explain student performance in general, especially cross-curricular competencies such as problem solving. PISA 2012 will cover a measure of students' openness for learning and exploration, plus a taxonomy of problem solving approaches based on vignettes that students comment upon, and a measure of test-taking strategies.

Design considerations: While family background and support would best be captured by the parent questionnaire, most learner characteristics have been and will continue to be assessed within the student questionnaire (especially since the parent questionnaire is likely to remain an international option in 2012). Two of the international options proposed for PISA 2012 provide additional information, namely the educational career and the ICT questionnaire. The strongest student-level predictors of learning (*i.e.* intellectual capacity and previous domain knowledge), however, cannot easily be measured in PISA. However, longitudinal extensions to PISA, which have been successfully applied in some countries, have been able to take predictors such as intellectual composition and prior knowledge into account to estimate appropriately the value added measures of school effects. Consideration should be given to longitudinal extensions as an option if policymakers want to be informed accurately of the effects of schools and schooling (see Section 5.5).

3. Learning conditions for mathematical literacy

PISA 2012 focuses on mathematics as the major domain of student assessment, with clear implications for the design of both student and institutional level background variables. In line with the educational taxonomy shown in Table 1 above at the student level, PISA aims to “portray important aspects of the affective domain, information about students’ experience with mathematics in and out of school (*e.g.* experience of different approaches to teaching mathematics, preferred ways of learning), motivation, interest in mathematics and engagement with mathematics” (EDU/PISA/GB(2008)28, p.35, par. 168). At the institutional level (*i.e.* classrooms and schools in Table 1 above), PISA aims to “portray important aspects of learning and instruction in mathematics, including an investigation of the relationship between learning and teaching strategies and performance as well as the relationship between school organisation, and structures and active student engagement with learning” (p.37, par. 175). These aims, in turn, are in line with the “Longer-term strategy for the development of PISA”, in which the PISA Governing Board as early as 2005 stated that “it would be conceivable to make the effectiveness and efficiency of educational processes the overarching theme for the PISA 2012 assessment”, because such an analysis “would lend itself particularly well to the subject area of mathematics, as the most school-bound subject covered by PISA” (EDU/PISA/GB (2005)21, p.17, Par. 67).

In the following sections, foundations for the study of various facets of mathematical literacy, including individual prerequisites and institutional conditions will be specified. Moreover, conceptual issues will be supplemented by recommendations for actual measures to support the development for the PISA 2012 context questionnaires.

3.1 *Mathematical literacy: A challenge for instruction and assessment*

Since mathematical competency is one of the most important prerequisites for success in the modern, technology-rich, and knowledge-driven world, it is given priority status by most education systems. Mathematical competencies have been and still are an essential domain in nearly every large scale educational assessment – be it national or international. High quality mathematics instruction has received strong attention by educational policymakers worldwide. However, the very meaning of high quality instruction has changed dramatically over the last twenty years (see Schoenfeld, 2006). This reform movement originates from many sources, from Dewey’s pragmatism, the late Russian psychologist Vygotsky’s notion of cognitive development and German “Reformpädagogik” (*i.e.* reform pedagogy) to modern constructivist epistemology and the theory of self-regulated learning. Based on these sources, non-mechanistic concepts of learning and teaching were invented, which eventually permeated the professional practices of mathematics teachers.

Probably the most influential document of recent years that has shaped the understanding of mathematical education and has been echoed in many parts of the world, the “Curriculum and evaluation standards for school mathematics” was issued in 1989 by a professional teacher organisation in the USA. This document sets out “five general goals for all students: (1) that they learn to value mathematics, (2) that they become confident in their ability to do mathematics, (3) that they become mathematical problem solvers, (4) that they learn to communicate mathematically, and (5) that they learn to reason mathematically” (NCTM 1989, p. 5). The NCTM standards mark a substantive shift in the way mathematics education is conceptualised, because general competencies for mathematical thinking substitute the content-oriented learning goals that had previously characterised the purposes of education. Mathematics education now aims at fostering mathematical thinking and its application to real world problems, well beyond declarative knowledge and procedural skills. Subsequent research elaborated and expanded on those competencies through both theoretical work and empirical studies (Niss, 2003; Blomhoj & Jensen, 2007). Most recently, similar developments have been further expressed through the Common Core State Standards Initiative (2010) for Mathematics in the USA. As well as standards for mathematical content, the CCSSI draft

document includes a set of standards for mathematical practice that in part build explicitly on the NCTM standards mentioned above.

The mathematics assessment framework for PISA has been built on this new tradition, and in fact for one decade, PISA has been a major force in promoting the thinking/modelling approach to mathematics education, strongly supported by policymakers as well as mathematics education experts worldwide. The PISA Assessment Framework for Mathematics (OECD, 2009c) does not just mirror the traditional content strands of school mathematics, such as algebra, geometry, or statistics. Rather, it stresses the “big ideas” that guide conceptual understanding, and it requires mathematical competencies well beyond technical knowledge and skill.

When PISA was conceptualised, the fundamental idea was to assess mathematical thinking in context. In the draft framework for PISA 2012 mathematical competence is defined as “an individual’s capacity to formulate, employ, and interpret mathematics in a variety of contexts. It includes reasoning mathematically and using mathematical concepts, procedures, facts, and tools to describe, explain, and predict phenomena. It assists individuals to recognise the role that mathematics plays in the world and to make the well-founded judgments and decisions needed by constructive, engaged and reflective citizens” (OECD, forthcoming). When essentially the same concept was elaborated and transformed into test items for earlier PISA surveys, the items were designed as “measures of the extent to which students presented with (problems mainly set in real-world situations) can activate their mathematical knowledge and competencies to solve such problems successfully” (OECD, 2004, p. 37). The draft framework for PISA 2012 (OECD, forthcoming) continues to base the assessment of mathematical competence for 15-year-olds on phases of the mathematical modelling process, and builds on a central set of mathematical capabilities, namely Communication, Mathematising, Representation, Reasoning and Argument, Devising Strategies, Using Symbols and Formal Operations, and Using Mathematical Tools. Also, items are developed in line with four broad content categories of secondary school mathematics, namely space and shape, change and relationships, quantity and uncertainty. This means the PISA mathematics survey for the foreseeable future will continue to go well beyond technical mathematical content knowledge and skills.

The PISA cognitive assessment thus covers three out of the five goals of modern mathematics education cited above, namely students’ ability for mathematical reasoning, problem solving, and communication. Both remaining goals, valuing mathematics and being confident in doing mathematics, are assessed as non-cognitive outcomes within the PISA student questionnaire, along with mathematics-related strategies. Besides being warranted as outcomes of mathematics education, these constructs will also assist in explaining differences in performance on the PISA mathematics test. The next section will provide the conceptual background and measures.

3.2 Outcome measures related to mathematical literacy: Strategies, beliefs, and motivation

Strategies and metacognition

Mathematics educators want to know not only which and how many tasks or problems students can solve, but also *how* they are working on these problems (*i.e.* which strategies they apply in studying mathematics, preparing for tests, or tackling/ approaching problems). Domain-specific strategies of learning and problem solving are both an important outcome and a prerequisite for mathematical learning – much like text-based learning strategies within reading education. Good knowledge and command of strategies allows for the transfer of mathematical literacy into new contexts and new tasks.

In the context of reading literacy, PISA has applied student self-assessment of three cognitive learning strategies, namely memorisation (*e.g.* learning key terms, repeated learning of material), elaboration (*e.g.* making connections to related areas, thinking about alternative solutions), and control (*i.e.*, meta-cognitive

strategies that involve planning, monitoring, and regulation). PISA 2003 adapted these scales to the domain of mathematics. The elaboration scale used several items that had been adopted from the domain of reading, like “When learning mathematics, I try to relate the work to things I have learned in other subjects”. In addition, the elaboration scale included specific items like “When I am solving mathematics problems I think of new ways of getting the answer” and “... I think of how the solution might be applied to other questions”. This modified elaboration scale proved to be a valid and strong predictor of mathematics performance across countries. At the same time memorisation does not seem to be identifiable as a separate and valid scale at all, and mixed results have been found for self-assessed control strategies (OECD, 2005a, p. 297; Vieluf, Lee, & Kyllonen 2009a, 2009b).

It is not by chance that the mathematics-specific elaboration items have been situated in the context of solving mathematics problems. International comparative research has shown that mathematics instruction all over the world is mostly defined by working on problems (Christiansen & Walther, 1986; Hiebert *et al.*, 2003; Stigler & Hiebert, 1999). When students practise individually, in small groups, or as a whole class, they are regularly assigned either to problems or routine tasks. Even when new content is introduced, teachers in many countries use example problems to work with the class through new solution methods. Thus, studying (for) mathematics or learning mathematics basically means solving assigned problems. Rather than learning strategies, as in the domain of reading, problem solving strategies are at the heart of mathematical literacy. Cognitive research in mathematics learning re-specified these as instantiations of what is now commonly called metacognition (Desoete and Veenman, 2006; Garofalo & Lester, 1985; Schneider and Artelt, forthcoming; Schoenfeld, 1992). In general, metacognition refers to the individual’s own awareness and consideration of his or her cognitive processes and strategies, as defined in the now classical paper by Flavell (1979).

Design considerations: To allow for trend analysis, we suggest that the Elaboration scale from PISA 2003 (a self-assessment of mathematical learning and problem solving strategies) be retained. However, as Schneider and Artelt (in press) point out, self-assessments of knowledge about and/or use of strategies are weak indicators of metacognition. In addition, indicators based on logfile data from the computer based cognitive skills tests, or indicators of declarative metacognitive knowledge similar to the PISA 2009 instrument which assessed metacognition in reading could be considered.

Motivation and intentions

In an era when many countries suffer from low enrolment in Science, Technology Engineering, and Mathematics (STEM) subjects, especially among females, the strengthening mathematics interest and motivation or more general positive attitudes towards mathematics has become a major policy issue. Also, evidence abounds of a positive relationship between mathematics-related attitudes and mathematics performance (see, *e.g.* the meta-analysis by Ma & Kishor, 1997).

Theories of student motivation usually distinguish between intrinsic and extrinsic motivation. Intrinsic motivation comes from rewards inherent to a task or activity itself. PISA 2003 assessed this type of motivation with regard to mathematics by means of a scale labelled, Interest In and Enjoyment of Mathematics (example item: “I am interested in the things I learn in mathematics”). The research literature suggests that intrinsic motivation has an effect on student engagement and time-on-task, learning activities, performance, and career choices, and it can be shaped by classroom instruction (Kunter, 2005; Rakoczy, Klieme and Pauli, 2008; Ryan and Deci, 2000). Extrinsic motivation, by contrast, comes from outside the subject, *i.e.* from expected rewards. It has been found to be an important predictor of course selection, career choice, and performance (Wigfield, Eccles, & Rodriguez, 1998). In PISA 2003, a scale for Instrumental Motivation to Learn Mathematics was applied. One of the items reads as follows: “Learning mathematics is worthwhile for me because it will improve my career options”.

More sophisticated models of behaviour regulation distinguish between more general dispositions like intrinsic and extrinsic motivation, and actual intentions. Intentions are closer to real decision-making, and are more relevant for assessing how students deal with mathematics in their present life and in their plans for the future. Such items are presumably less biased by culture-specific response styles or peer-group norms. Short-term intentions may be assessed by a scale with items like “I will try to work hard to make sure I learn mathematics” (Lipnevich *et al.*, in press). Long-term intentions or future orientations may be captured by adapting a scale that was introduced for science-related studies in PISA 2006, with items like “I would like to work in a career involving mathematics”. The latter scale would allow the examination of whether education systems have a chance to recruit high achieving students for career paths which make use of these competencies. Also, issues of equity in access to STEM careers can be investigated.

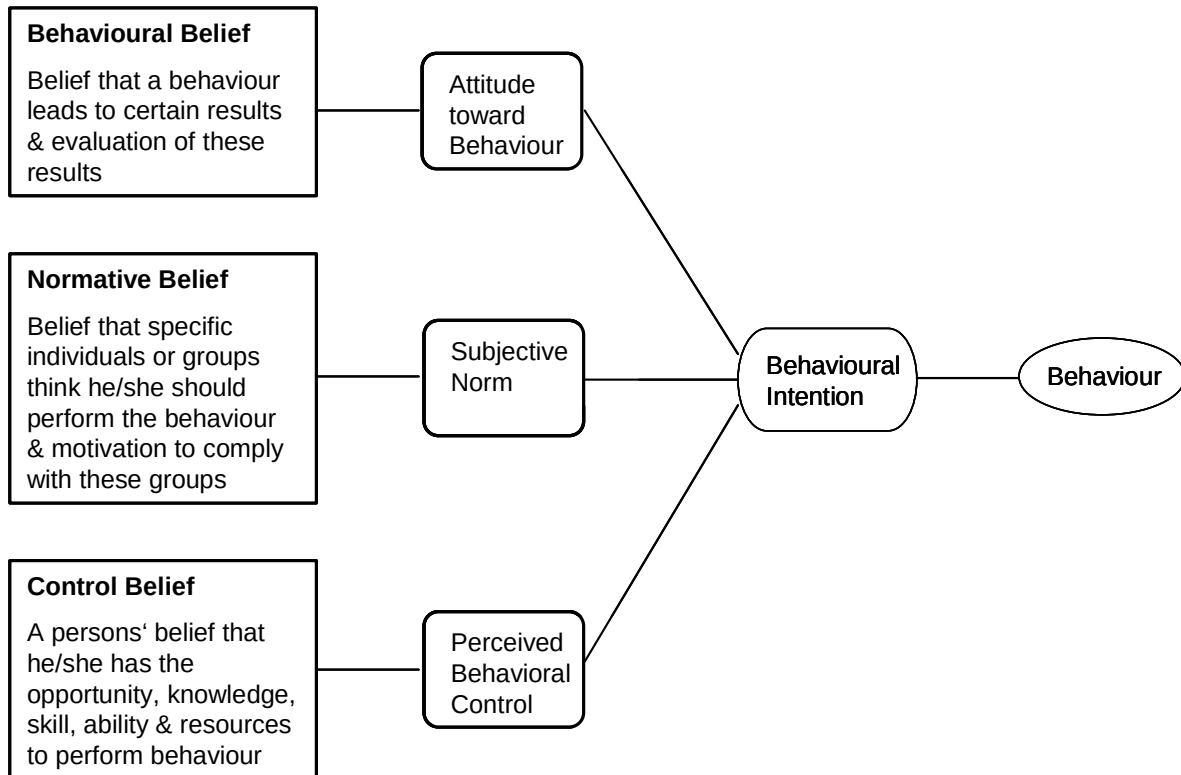
Design considerations: These four scales, namely Interest In and Enjoyment of Mathematics, Instrumental Motivation to Learn Mathematics, Short-Term Intentions, and Long Term Intentions will suffice for addressing policy-relevant issues in student motivation for mathematics learning. Applying six attitudinal scales within the student questionnaire plus several embedded measures within the cognitive test, as in PISA 2006, would unnecessarily overburden students. However, innovative approaches will be tried out to reduce cultural bias and assure the cross-cultural comparability of results (see Section 5.3 below).

Self-related beliefs and planned behaviour

How students think and feel about themselves is an important predictor for how they act and decide when challenged by tasks and situations (Bandura, 1977). While self-efficacy is defined as one’s conviction or belief about his or her capability to cope with certain tasks, self-concept is defined in a broader sense as the overall perception of one’s personal attributes based on continuous self-evaluation. Research into mathematics education found that subject-specific self-efficacy predicts mathematics grades, mathematical problem solving, interests, and – once again – career choices (Britner & Pajares, 2006; Pajares & Miller 1997; Turner, Steward, & Lapan, 2004). PISA 2003 assessed both mathematics-related self-concept (example item: “I learn mathematics quickly”) and self-efficacy in relation to mathematical tasks (see Lee, 2009; OECD, 2005a). More specifically, students were confronted with eight tasks, ranging in difficulty from straight-forward technical procedures like solving the equation $3x + 5 = 17$ to mathematical modelling tasks like calculating the petrol consumption rate of a car (OECD, 2005a, p. 292), and asked “How confident do you feel about having to do the following calculations?”. Both self-concept and self-efficacy turned out to be cross-culturally equivalent, strong predictors of student performance (Vieluf, Lee & Kyllonen, 2009a, 2009b).

How confident students are about their ability to solve mathematical tasks, as well as how students value mathematics, are highly relevant factors in predicting or explaining student behaviour with regard to mathematics, *e.g.* course-taking and career decisions. A number of Expectancy Value models both in psychology and in economics have been proposed to integrate both aspects of decision-making. One such model is Ajzen’s (1991) theory of planned behaviour, which states that volitional behaviour is determined by specific attitudes and subjective norms (= value component) plus perceived behavioural control (= expectancy component). (For an application to mathematics in cross-cultural research, see Lipnevich *et al.*, in press.) According to Ajzen (1991), by manipulating these predictors, the chance that the person will intend to do a desired action can be increased. This, in turn, will increase the likelihood of the behaviour actually occurring (see Figure 1).

The theory of planned behaviour0



In the current context, students’ attitudes, perceptions of control, and subjective norms, may predict their desire to spend time on mathematics homework, to ask questions in class, or to engage in relaxation exercises to counteract anxiety, and thus enhance their mathematics performance.

Design considerations: Application of the full Ajzen model will substantially increase the analytical power of the study for explaining mathematics-related effort, mathematics-related student behaviour, and indirectly, if possible, student outcomes. However, because of the larger costs and efforts that are needed for measuring all facets of the model, this extension would be given to a subset of students only, based on the suggested rotation design (see Section 4.3 below). Within the student questionnaire, self-concept and self-efficacy should be used again and a measure of mathematics anxiety retained. However, as its cross-cultural validity and incremental predictive power are questionable (Lee, 2009; Vieluf, Lee & Kyllonen, 2009a, 2000b), new item formats should be tried out for this scale.

3.3 Opportunity to learn and quality of instruction: Assessing the learning environment

PISA intends to cover learning conditions for the major domain in a particular cycle for two reasons: First, PISA helps to describe and compare learning environments for 15-year-old students in different countries, based on measures that have been shown to be important indicators of quality instruction elsewhere in the research literature. Second, those measures can help to explain student outcomes – not in every country, but in those countries where most students have been attending the school in which testing occurs for at least two years. This is the case in two thirds of countries that participated in PISA 2006.

Learning environments that help students to acquire mathematical literacy – as defined in PISA – have been described in the research literature as learning in context (Schoenfeld, 2006), discussion-based

teaching (Boaler & Greeno, 2000; Stein *et al.*, 2008), and they include mathematical modelling (Blum & Leiss, 2005) as an essential activity. In previous, seminal work, Schoenfeld (1987, 1992) called for a mathematics culture that allows for meaningful learning, makes connections between mathematical concepts, and links these constructs to students' everyday life – as compared with traditional instruction that aims at mastering formulas, algorithms, definitions, and other technical content. The “current vision of effective mathematics teaching”, as Stein *et al.* (2008) refer to it, includes the following: “Students are presented with more realistic and complex mathematical problems, use each other as resources for working through those problems, and then share their strategies and solutions in whole-class discussions that are orchestrated by the teacher” (p. 315).

As PISA intends to cover major aspects of the learning environment for mathematics within its questionnaire design, this vision of modern mathematical education becomes a challenge. In order to describe teaching and learning processes that are conducive for mathematical literacy, it is insufficient to ask about content coverage: First, because mathematical literacy as defined by PISA is a yield measure, dependent on fundamental mathematical abilities and processes, and hence less dependent on specific content elements that have been taught at school. Second, the quality of mathematics instruction is not determined by content but rather, by clarity and structure, support for students, and cognitive challenge (see Section 2.4 above). Therefore, PISA needs to re-define the notion of opportunity to learn (OTL) that has been an important construct characterising learning environments in other international student assessments. In addition, within the PISA design, OTL data obtained from students are examined at the school level.

Opportunity to learn – Coverage of content categories and problem types

The notion of opportunity to learn was introduced by John Carroll in the early 1960s, and was initially meant to indicate whether students had sufficient time and received adequate instruction to learn (Carroll, 1963; cf. Abedi *et al.*, 2006). It has since been an important concept in international student assessments (Husén, 1967; Schmidt & McKnight, 1995; Schmidt *et al.*, 2001), and has been shown to be strongly related to student performance, especially in cross-country comparisons (Schmidt & Maier, 2009, pp. 552-556). At the same time, the meaning of the construct became much broader. Stevens (1993, pp. 233-234) already identified four kinds of OTL variables most prevalent in research:

- **Content Coverage Variables:** These variables measure whether or not students cover the curriculum for a particular grade level or subject matter.
- **Content Exposure Variables:** These variables take into consideration the time allowed for and devoted to instruction (time-on-task) and the depth of the teaching provided.
- **Content Emphasis Variables:** These are variables that influence which topics within the curriculum are selected for emphasis and which students are selected to receive instruction emphasising lower order skills (*i.e.* rote memorisation) or higher order skills (*i.e.* critical problem solving).
- **Quality of Instructional Delivery Variables:** These variables reveal how classroom teaching practices (*i.e.* presentation of lessons) affect students' academic performance.

Thus, for certain authors, OTL has become more or less a synonym for the quality of instruction experienced by the student. Schmidt and Maier (2009), however, in their review argue that opportunity to learn (OTL) is a rather uncomplicated concept: “What students learn in school is related to what is taught” (p. 541), and they intentionally focus on OTL “in the narrowest sense: Student’s content exposure” (p. 542).

Schmidt and Maier acknowledge that although OTL may be a straightforward construct, it is quite difficult to measure. In order to explain differences in the achieved curriculum, traditionally, teachers and/or students have been asked whether and how certain curricular content has been realised in instruction (the implemented curriculum), sometimes using logs (Rowan *et al.*, 2004). In addition, curriculum experts have been asked whether and how content elements have been covered within curricular documents like syllabuses, textbooks, and standards (the intended curriculum). From these raw data, various indicators have been extracted: In many cases, the content taught has been judged twofold, in terms of topic and level of demand, while at the system level indices for coherence, rigor, and focus have been derived (Schmidt & Maier, 2009).

Design considerations: In PISA, the measurement of OTL has to be modified from approaches used in other studies as the mathematics assessment is not framed according to content elements, but refers to seven fundamental mathematical abilities and four content categories. In PISA, the operationalisation of OTL will be built mainly on student judgements.

PISA 2012 will aim to identify country (and probably school) level profiles in learning opportunities. Students will be confronted with carefully crafted mathematics tasks – some representing mathematical abilities and content categories as mentioned in the PISA Mathematics Framework, some representing more traditional tasks asking for procedural and declarative knowledge. Following each of those items, students will be asked to judge whether and how often they have seen similar tasks in their mathematics lessons and in previous assessments. Thus, it will be possible – aggregated at the country, but possibly also at the school level – to measure learning opportunities in a way that allows for differentiation between types of problems and content.

In addition, students will be asked to estimate their familiarity with certain mathematical concepts. This measure, as well as the self-efficacy scale described in Section 3.2, may be used as proximal indicators for opportunity to learn.

Learning time

Following Carroll (1963), numerous studies have shown that learning time is a major predictor of student outcomes in many subjects, including mathematics. PISA 2012 will apply a measure that was already used in 2003, to ensure trend analysis. Students were asked: “On average, how many hours do you spend each week on the following?” with items for homework, remedial and enrichment classes, work with a tutor, and out of school classes. In addition, the weekly number and duration of mathematics lessons will be gauged in both the student and the school questionnaires.

Quality of instruction

As mentioned in Section 2.4, research on secondary school mathematics instruction suggests that (a) structure and efficient classroom management, (b) teacher support and student orientation, and (c) challenge and cognitive activation have to be addressed as basic dimensions of instructional quality. Two of these dimensions are covered by existing PISA 2003 scales, namely Disciplinary Climate in the Mathematics Classroom and Teacher Support in the Mathematics Classroom, both in the student questionnaire.

The international PISA 2003 report (OECD, 2004) showed that the construct Disciplinary Climate in the Mathematics Classroom was strongly associated with mathematical literacy, while other variables – such as class size, mathematical activities offered at school level, and avoidance of ability grouping – had no substantial effect once the socio-economic status was taken into account. These findings are in line with previous research and with the theory outlined in Section 2.4, which predicts that well-structured, efficient

classroom management is a prerequisite for student learning. However, the PISA 2003 report did not study the relationship between learning environment and affective outcomes, such as students' interest and enjoyment in mathematics. Meanwhile, Vieluf, Lee and Kyllonen (2009a) have shown that teacher support is positively linked to students' interest in mathematics, also after controlling for socio-economic status.

Challenge and cognitive activation, however, are very difficult to assess, although this dimension is crucial for fostering mathematical literacy. Similar concepts, emphasising particularly the importance of a demanding orchestration of teaching the mathematical subject matter which gives learners vast opportunities to develop and practice competencies, are broadly discussed in mathematics education (see, *e.g.* Blum & Leiss, 2007). However, none of the indicators used in PISA 2003 could capture this dimension. Several approaches have been used successfully in national studies and are proposed for the PISA 2012 field trial:

- As homework is an almost universal element of mathematics instruction, the way teachers deal with homework during lessons can be used as an indicator for challenge. Rakoczy, Buff and Lipowsky (2005) developed a scale capturing the process-oriented dealing with homework, with sample items "My mathematics teacher is eager to learn how we solved problems in our homework assignment" and "When talking about homework results, we try to understand and correct student errors". (Note that the scale does not address homework assignments per se, or time used for homework, but the quality of classroom discourse about homework results).
- Students can also be asked about the type of items they usually encounter when practicing in their mathematics classes, *e.g.* "We usually have to think for a while in order to solve the problems we are assigned by our mathematics teacher". (Scale "cognitive activating tasks" by Baumert *et al.*, 2008, used in PISA 2003/Germany.)

Teaching practices and students' mathematical activities

A behaviour-oriented version of the triarchic model of instructional quality has been implemented in the OECD TALIS study by asking teachers how often they implemented each of 13 given practices in their teaching:

- Structuring practices (5 items): *e.g.* "I explicitly state learning goals." Other items include summary of former lessons, homework review, checking the exercise book, and checking student understanding during classroom talk by questioning students.
- Student-oriented practices: (4 items): *E.g.* "Students work in small groups to come up with a joint solution to a problem or task." Other items include ability grouping, student self-evaluation and student participation in classroom planning.
- Enhanced activities (4 items): *e.g.*: "Students work on projects that require at least one week to complete." Other items include making a product, writing an essay, and debating arguments.

Based on TALIS main study data from 23 countries, it has been shown that (a) the three dimensions can be differentiated across countries (*i.e.* the triarchic model has some cross-cultural validity), (b) structuring practices, as hypothesised, are associated with higher levels of classroom discipline (as perceived by teachers), and (c) participation in professional development as well as teaching high ability classes raises the frequency of using these practices. Mathematics and science teachers report less student orientation and less frequent use of enhanced activities than did teachers who teach other subjects (OECD, 2009b).

Design considerations: The TALIS teaching practices scales could be adapted for use in PISA 2012 within a brief add-on to the TALIS teacher questionnaire (see document EDU/PISA/GB(2010)10). Items for

enhanced activities would have to be reframed for mathematics, with examples such as: “Using mathematics to solve an everyday problem”, “Connecting mathematics to other subjects”, “Inducing rules from mathematical patterns”, “Elaborating a proof of a mathematical theorem”, “Discussing examples and counterexamples”, “Using multiple graphical representations for a mathematical problem”, “Comparing different solution methods, or multiple solutions for the same task”.

While such a link to TALIS would be feasible in only some PISA countries, all of the students could be asked how often they actively participate in such activities.

3.4 School- and system-level support for teaching and learning of mathematical literacy

Most of the school- and system-level variables that have been discussed in Sections 2.2 and 2.3 are likely to be linked to mathematics performance and/or mathematics-related attitudes. Input and process indicators specific to mathematics can be obtained by aggregating student responses about OTL and quality of instruction (Section 3.3) to higher levels.

However, a number of mathematics-specific variables at the school or the system level will directly shape conditions for teaching and learning mathematical literacy. Therefore, special emphasis may be given:

At the system level to...

- a) reform initiatives, school improvement activities, and professional development programs targeted at mathematics,
- b) the role of mathematics in student admission, tracking, assessment and promotion as well as in school evaluation,
- c) the intended instruction time for mathematics per year,

At the school level to...

- d) mathematics-related course offerings (i.e. remedial lessons, enrichment activities, tutoring sessions),
- e) qualification of the mathematics teaching staff,
- f) collaboration among mathematics teaching staff .

Design considerations: PISA is asking the school principal about all of these issues. The data may also be aggregated to the system level.

4. Specifying the questionnaire design for PISA 2012

4.1. Distributing content across PISA cycles: An overarching design for general, domain-specific, and thematic extension variables

Some of the relevant factors in understanding student performance, attitudes, and behaviours, and the functioning of education systems are straightforward (such as demographic variables, previous educational career choices, instructional time, and class size), some have been well established in previous PISA cycles (such as student socio-economic status, cognitive strategies, school-level decision-making), while others have proven to be less easily addressed within the PISA design (*e.g.* accountability policies at the system level, teacher variables, aspects of the classroom learning environment, or out-of-school activities).

Choosing among the many variables that might be incorporated into the design is a complex process, directed by the priorities that countries have set for the study, but also informed by educational research, as outlined in the previous sections.

Given that PISA has multiple purposes, and addresses multiple outcomes of student learning, and considering also that trend information is widely required, the Context Questionnaires must cover:

(I) General variables (for all cycles)

General input variables:

- Student-level inputs (grade, gender, socio-economic background= parental education and occupation/family wealth/educational resources/cultural possessions, migration data = immigration status / heritage language/ age on arrival in country, family support)
- School-level contexts and inputs (community size, resources, qualifications of teaching staff)

General process variables:

- School-level processes (decision-making, admission policies, assessment and evaluation policies, professional development, teacher engagement/morale, teacher-student relations, parental involvement)
- Instructional processes (learning time, disciplinary climate, teacher support)

General outcome variables:

- General non-cognitive outcomes – Commitment to learning (behavioural: truancy; personal goal: educational aspirations; motivational: learning engagement, affective: sense of belonging)

(II) Domain-specific trend variables (for major domain only, included every 9 years)

- Domain-specific non-cognitive outcome variables (strategies and metacognition, domain-related beliefs, self-related beliefs, motivation)
- Domain-specific processes variables (Opportunity to learn, instructional quality, system- and school-level support)

(III) Thematic extension variables (extensions within individual cycles)

- International options (*e.g.* in 2012, educational career/second language learners; ICT literacy)
- Context variables for additional domains (*e.g.* ICT-related experiences relevant for computer-based problem solving)
- Descriptive and explanatory variables for specific reports (*e.g.* in 2012: mathematics-related motivations and intentions)
- Malleable variables at the school level (*e.g.* tracking policies, teacher certification) that are specifically selected for descriptive purposes or for causal inference.

(IV) System-level data, gained from INES or system-level questionnaire

- Output of educational institutions (*e.g.* certificates)
- Financial and human resources invested into education
- Access to and participation in education
- Learning environment and organisation of schools

In previous PISA cycles – especially in PISA 2006 – the questionnaire design largely neglected constructs that were not related to the major domain. Thus, the set of general variables (I) was sometimes restricted to include student inputs only. Student, and to some extent also school, questionnaires focused on processes and non-cognitive outcomes that were domain-specific. However, this design is inappropriate for measuring trends.

Finding an appropriate balance between (I), (II), (III), and (IV) is crucial for the overarching design of PISA questionnaires, and for the long term success of the PISA program.

In order to establish valid and reliable trends at the country level, it is important to implement a constant set of general variables in all cycles both for the calculation of proficiency estimates and as major reporting variables. Thus, these context and input background variables should not change.

Policy-makers do not only need trend information on student performance. Education systems serve other goals as well. Thus, policy-makers ask for information on domain-independent, non-cognitive or behavioural outcomes, namely educational aspirations (occupation expected at age 30), school commitment (sense of belonging), truancy, and learning motivation (*e.g.* effort and perseverance) – all covered in previous PISA cycles. So far, PISA authors have been reluctant to use these variables, fearing their insufficient cross-cultural comparability. However, as trends are available over an increasing number of cycles, the focus now is on change rates within countries, rather than on cross-sectional comparisons of status. For example, whether and to what extent school commitment is increasing or decreasing, will be a relevant indicator within countries. Also, cross-cultural validity will receive more attention (see Section 5.2 below).

Two process scales, namely Disciplinary Climate in the Classroom and Teacher Support in the Classroom, are related to student performance and student interest, respectively (see Klieme & Rakoczy, 2003; Vieluf, Lee and Kyllonen, 2009a). We thus recommend the use of these scales, in addition to learning time, as general process variables, to be retained in future cycles. Instead of changing the wording each time, however (*i.e.* instead of applying them to the major domain only), the wording should be kept stable and domain-independent, in order to allow for change models to be applied.

Depending on the major domain of cognitive assessment, additional measures shall be taken into account (domain-specific trend variables). And finally, each cycle will have its specific foci, either in the form of international options, or as extensions to compulsory questionnaires.

4.2 Exploration of measures from PISA 2003

The following aspects of the learning environment for mathematical literacy were captured in the 2003 data set:

- Class size (student questionnaire)

- Disciplinary Climate in the Mathematics Classroom (student questionnaire)
- Teacher Support in the Mathematics Classroom (student questionnaire)
- Time Spent on Mathematics Instruction at School and at Home (student questionnaire)
- Frequency of Ability Grouping for Mathematics (school questionnaire)
- School activities that promote student engagement with mathematics, such as Participation in Mathematics Competitions (school questionnaire)
- Characteristics of the mathematics teaching staff, namely Level of Qualification, Teacher Consensus, and Evaluation Policies (school questionnaire).

Several mathematics-related outcome variables were assessed as well: Learning strategies, Interest in and Enjoyment of Mathematics, Instrumental Motivation, Self-efficacy, Self-concept, and Mathematics Anxiety.

In order to deliver trend information on non-cognitive outcomes and mathematics-related context/process variables, PISA 2012 will have to retain as many variables that were used in 2003 as possible, unless they were shown not to work cross-culturally or not to account for differences in outcomes. Table 2 gives an overview of all questions from the PISA 2003 student and school questionnaires. The first row identifies the individual inputs – student demographic and family background variables – which have remained stable over the cycles⁴. These input variables are used to study the distribution of educational resources and outcomes within countries, and they are treated as control variables in analytical models.

The second row of the table contains school input variables such as the public versus private status of the school and its financial, staff, and other resources, including the size and composition of the student population. These variables are inputs at the school level but budgetary decisions, overall changes in the program structure, and privatisation policies are at least partly controlled at the country or system level.

Based on the technical characteristics documented in the PISA 2003 technical report (OECD, 2005a), taking also into account multilevel analyses of predictive power (Vieluf, Lee and Kyllonen, 2009a), cross-cultural validity (Vieluf, Lee and Kyllonen, 2009b) and a review of the use of information from the school questionnaire in previous PISA cycles (Hersbach & Lietz, 2010), a set of variables that will be retained as school input variables has been selected.

At the individual student level, results of the predictive power analyses have shown that the link between learning preferences and mathematics performance were relatively weak: The link for cooperative learning was particularly weak. In similar analyses of PISA 2000 data (Jehanghir & Glas, 2007), the relationship to reading performance in PISA 2000 of cooperative learning was shown to be non-significant and that of competitive learning to be rather small. Hence, these two scales might be deleted. Among the three strategy scales, memorisation as a domain-specific learning strategy was not related to mathematics performance. In addition, cross-cultural invariance – even at the metric level – could not be established for this scale. Therefore, it might also be removed. In contrast, elaboration was strongly linked to performance and was acceptable in terms of cross-cultural invariance (see Section 3.2). Finally, the attitudes towards school scale might also be considered for deletion as a consequence of its weak link to student performance and lack of cross-cultural invariance.

⁴ . Variables that were used in 2003 only, or dropped later, include family structure, educational career, and reason for attending school.

At the school level, relatively strong effects were found for Sense of Belonging, Time Spent on Mathematics Homework and Disciplinary Climate in the Mathematics Classroom as process predictors of mathematics performance as well as mathematics interest. Consequently, these variables will be retained whereby Disciplinary Climate in the Mathematics Classroom demonstrates the strongest (scalar) level of invariance, which allows for comparison of mean scores between countries.

Self-efficacy, Self-concept, and both motivation scales have scalar invariance, allowing use of these variables as non-cognitive, mathematics-related outcome indicators, establishing trend, as well as cross-national comparative analyses. However, alternative measures will be tried out that may further improve cross-cultural comparability.

A classification of PISA 2003 questionnaire

	Student questionnaire	School questionnaire
<i>Antecedents</i> (Student background and school context)	Age, Gender, Immigration background Heritage language, Parental occupation, Parental education, Family wealth and possessions (= 4 indices), Family structure <u>Educational career, Current grade, Study program, Reasons for attending, Study time outside school, Expected educational level</u>	Community size
School input (controlled at the system/ school level, depending on governance structure)		Enrolment per gender, Proportion of 2nd language learners. Public/private ownership, Grades covered, Schedules per program, Quantity of teaching staff, Availability of computers, Financial resources, School resources (3 scales)
<i>General processes</i> (controlled at school, sometimes at system level)	Student-teacher-relations	Decision-making/different sorts of autonomy, Admission policies, Assessment policies, Options for 2nd language learners, Teacher morale, Negative teacher behaviours, Teacher participation,
<i>Mathematics processes</i> (controlled at instruction or school level)	Instructional time, Time spent on mathematics homework Class size Disciplinary climate, Teacher support	Ability grouping in mathematics, Mathematics activities Mathematics teaching staff: Qualification + evaluation policies; Mathematics teacher consensus
<i>General non-cognitive outcomes</i>	<u>Attitudes towards school, Truancy,</u> Sense of belonging to school	Negative student behaviours <u>Student morale</u> Grade repetition
<i>Mathematics non-cognitive outcomes</i>	Interest in and enjoyment of mathematics, Instrumental motivation, Self-efficacy, Self-concept, Mathematics anxiety, strategies (control/ elaboration/ memorisation), <u>learning preferences: Co-operative vs. competitive</u>	

Underlined constructs might be considered for deletion

Bold constructs should be kept for 2012

4.3 *Rotation design for extended student questionnaires and computer-based delivery*

Covering all of the general, domain-specific trend, and thematic extension variables would definitely go beyond the scope of time and space allocated for context questionnaires in PISA. This potential problem especially holds true for the student questionnaire, which is used, by means of data aggregation, to generate information for all levels of the study design (*i.e.* from the individual student up to the system level). The school questionnaire is also in danger of being extended, facing an increased risk of non-responses. To allow PISA 2012, as well as future cycles, to serve the intended multiple purposes, and to prevent an increase in missing data, new methods to reduce response burden are needed. To this end, two technical innovations are proposed for PISA 2012.

Rotation

The consortium suggests a rotation design for the student questionnaire, similar to the design for cognitive items, which makes use of item packages distributed over a number of different booklets. Each student is allocated one of these booklets, and thus receives a limited number of cognitive tasks, while the booklets taken together cover a large “universe” of items from various domains. Booklets are randomly rotated among test-takers within each test site (*i.e.* school), hence the name of the design. Such rotation designs are common in large-scale assessments, both in national and international surveys.

Applying a rotated design to questionnaires is relatively new to educational measurement, though it is a well-established technique in other kinds of surveys. The implications of such a design for (a) cognitive proficiency estimates, (b) the international report and the reporting of trends, (c) further analyses, (d) documentation and structure of the international database, and (e) logistics” have been discussed elsewhere (Berezner & Lietz, 2009). Taking the existing experience into account, and based on analyses of PISA2006 data, it is recommended to use a rotated design for the PISA 2012 student questionnaire. In this way, it is estimated that the amount of material covered can be increased by one third.

The common, non-rotating part of the student questionnaire, should at least comprise the student-level general input variables (see Section 4.1 above). This is based on methodological and reporting requirements, theoretical work and analyses of 2006 PISA data (Berezner & Lietz, 2009).

In addition, the common part is suggested to contain constructs that are needed to explain differential effects on major minority groups. From a policy point of view, this includes all variables that may help to explain differences between students with and without immigration background. Furthermore, process variables that are based on the student questionnaire, but used on an aggregated level to describe the learning environment, shall be included in the common part, because otherwise the aggregated measure will not be reliable.

In effect, this means that most of the general outcome variables, the domain-specific outcome variables and the thematic extensions variables might go into either the rotated parts of the student questionnaire, or into international options, while most of the process variables will remain in the common, non-rotating part.

Computer-based delivery

Computer-based delivery allows for more flexibility, adaptivity, and efficiency in administering and scoring questionnaires. For example, filter questions can be easily applied to guide respondents through the

questionnaire, inconsistencies can be checked online, and no manual data entry has to occur. Thus, in the long run, computer-based delivery might also help to improve administration of the instruments.

PISA 2012 will use computer-based delivery for the school questionnaire. Subsequent cycles may also adapt this technology for the student questionnaire and international questionnaire options.

5. Assuring validity evidence, explanatory power, and policy relevance of the PISA design

In order to enhance validity and cross-cultural comparability, to define value-added indicators at all levels, and to enhance the explanatory power of the study, the QEG makes several recommendations regarding the PISA study design and analysis plan. Strengthening the design from a research methodology point of view will ultimately allow for a greater number, and increasingly sophisticated series, of policy-relevant statements to be supported by PISA data.

Current design and standard reporting methods allow, for example, the amount of remedial education offered by schools (as reported by school principals), and the amount of out-of-school tuition (as reported by students and/or parents) to be documented. Policymakers may ask: Do schools differ in the amount of remedial education offered? Do students from various social and ethnic backgrounds and various tracks (as identified in the PISA sampling framework) have equal access to remedial education? Is there positive discrimination in favour of disadvantaged students? Does private tuition compensate for lower opportunities for school-based remedial education? How do these relationships develop over time (across PISA cycles)? How does my country behave with regard to these figures compared to other countries?

More sophisticated modelling, however, is needed to answer the following policy questions, and more caution is needed in interpreting related findings: Does paid work prevent students from participating in remedial education and/or tutoring? Does individual participation raise student performance? Does offering remedial education raise mean performance at the school level, reduce variation in student performance within schools, or reduce the social gradient (*i.e.* the association between socio-economic status and performance) within schools? These are hypotheses about the impact of one treatment (*i.e.* paid work, tutoring, and offering remedial education) on another variable (*i.e.* participation in these activities, student performance). Testing such hypotheses requires some kind of causal inference. The validity of these inferences depends on the availability of control variables that may explain selection and self-selection into these treatments (Section 1.2). Analyses should be undertaken country by country.

Allocation (or self-selection) for remedial activities is assumingly driven by recommendation or pressure from teachers and parents, and influenced by student grades, gender, socio-economic background, and the educational program the student attends. In order to control for these conditions and to cover treatment variables, all students should be questioned (a) about the kind and duration of remedial mathematics they have received during the present school year, (b) about the grades that they received at the end of the previous school year, and (c) whether, at the end of the previous school year, teachers and/or parents recommended remedial training. The QEG recommends including these questions in the student questionnaire as a prerequisite for causal inference on this matter. Similar treatment effects could be studied in future cycles.

At the school level, the effect of a school's absenteeism policy and its implementation on student truancy, sense of belonging, and performance should be tested. Among other aspects, every school principal should be questioned on whether absenteeism was perceived as a severe problem by school authorities three years ago, and what kind of measures schools have been put in place since that point in time.

5.1 *How to model the effects of context variables on student learning and performance*

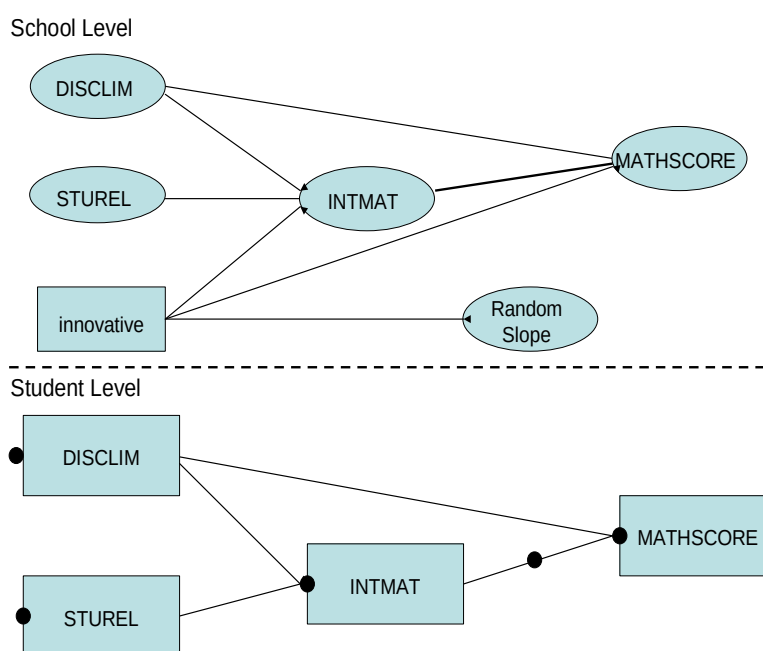
Previous PISA cycles and school effectiveness research more generally have studied the effects of school- and student-level processes on student performance and student motivation (*e.g.* interest in mathematics), controlling for inputs such as gender, socio-economic status, and migration status (at the individual level) and the respective aggregates (at the school level). By including such control variables, erroneous “false positive” conclusions about relationships between two variables that are in fact only due to a third variable can be avoided. For example, in many countries, a large proportion of the difference in performance between immigrants and non-immigrants can be explained by the fact that immigrants often have lower socio-economic status and therefore less social and cultural capital. Ignoring the difference in socio-economic status would lead to an overestimation of the effect of the migration background. Moreover, PISA has used multilevel modelling techniques like hierarchical linear modelling (HLM) to account for the hierarchical structure of data with students nested within schools and schools nested within countries. An underestimation of standard errors that can result from a violation of the assumption of independent observations within the sample can thus be circumvented, because students within schools are likely to be more similar regarding several characteristics (*e.g.* Raudenbush & Bryk, 2002; Snijders & Bosker, 1999). Multilevel models further allow for the examination of effects at different levels; that is, the effects of school level and individual level input variables and processes.

However, research in educational effectiveness shows that – when modelling relationships between input , processes and outcomes- three additional observations from previous research should be accounted for namely that effects are often (1) nonlinear, (2) moderated by other variables, and (3) that they may be indirect, or mediated by third variables.

- Non-linear effects refer to the fact that more is not always better, but rather sometimes a medium value is optimal. For example, a study by Monk (1994) suggests that the relationship between teachers’ subject knowledge and student performance may be curvilinear: On the one hand teachers are obviously not able to help students to understand content they are themselves not familiar with, but on the other hand a teacher with very sophisticated knowledge may have difficulties communicating knowledge in a way that is easily understood by students.
- In line with the theory of differential effectiveness (*e.g.* Kyriakides & Tsangaridou, 2004), it is important to acknowledge that relationships between variables may not be similar in different sub-groups. For example, there is some evidence that students from diverse social backgrounds may benefit from different instructional techniques (*e.g.* Brophy, 1992; Walberg, 1986). Based on a constructivist understanding of student learning, current educational theory assumes that student learning is largely dependent on self-regulated processes, which are moderated by school, classroom, and teacher factors. Modelling such differences requires the examination of interaction/moderation effects.
- Finally, it is reasonable to assume that not all effects on student outcomes are direct. Comparatively weaker effects on student outcomes are often found for policies at the school and system level, as compared to student background variables and classroom processes (*e.g.* Wang, Haertel, & Walberg, 1993). This may, in part, be due to the fact that the former variables do not exert a direct effect on students, but are rather related to school or classroom processes, which in turn have an effect on student performance. Moreover, school level variables such as school climate, shared values and norms, or procedures to deal with behavioural problems, may have a direct effect on non-cognitive outcomes (*e.g.* learning motivation, academic aspirations) and student behaviour (*e.g.* truancy, violence), while school effects on student performance and other subject-related outcomes (*e.g.* interest and self-efficacy beliefs) will most probably be mediated by teaching and learning within classrooms.

In some cases (1), (2), and (3) may apply simultaneously, for example when the effects of classroom climate and student-teacher relations on student performance are mediated by student interest. However, the effect of mathematics interest on student performance is stronger in schools where teachers use innovative, cognitively activating teaching practices, than it is in schools with a more traditional approach to teaching. (This model is exemplified in Figure 2. It is partly based on results reported from a bi-national study of mathematics teaching by Klieme, Pauli, & Reusser, 2009 and Lipowsky *et al.* 2009). A non-linear effect may be included when it can be shown that neither a too lenient nor a too strict classroom climate is optimal for enhancing student interest or performance. (For a more complex example based on analysis of one country's PISA 2003 data see Kaplan 2009b; Kaplan, Kim, & Kim, 2009)

An example of mediated moderation⁰



Notes:

DISCLIM – Disciplinary climate during mathematics lessons

STUREL – Student-teacher relations

INNOVATIVE – Innovative instructional strategies

INTMAT – Interest in and enjoyment of mathematics

MATHSCORE – Performance in mathematics

Moderation, mediation, and non-linear effects, can be modelled with traditional HLM techniques. However, for modelling mediating processes the application of structural equation modelling (SEM) is more convenient (see *e.g.* Kaplan, 2009b). The main advantage of multi-level SEM over multilevel regression analysis is that SEM uses latent instead of manifest variables, overcoming the biasing effects of measurement error at the student level (*e.g.* Muthén, 2002), but also for measures that are aggregated to the school level to model contextual effects (Lüdtke *et al.*, 2008). Thus, using SEM models would improve the validity of conclusions about associations between different aspects of education systems. Using multilevel SEM also allows for more flexible combinations of mediation, moderation, and non-linear effects at different levels, as well as the analysis of categorical manifest and latent variables. For example, multiple-group SEM could be used to examine whether the effect of classroom disciplinary climate on student

performance is similar across countries, or whether this variable may be more important in countries in which teachers have less authority. Latent class models could be applied for testing the hypothesis that it is a combination of different learning strategies rather than a high frequency of any one learning strategy that is related to student performance. Mixture models would allow the identification of groups of students within countries where relations between disciplinary climate and performance differ (*e.g.* Muthén, 2002). Such models may better reflect the complex reality of educational processes and their antecedents and consequences.

In short, using multilevel SEM models would help policymakers and professionals to gain a more comprehensive understanding of relationships among (a) learning opportunities provided at the school level, (b) individual use of those opportunities (in terms of quantity and quality of individual learning activities), and (c) student outcomes, both cognitive and motivational. (As an example, see Kaplan, Kim & Kim, 2009, based on the analysis of one country's PISA 2003 data.)

In summary, a simple production function approach to the study of educational outcomes is not appropriate from an educational point of view. System- and school-level as well as policy interventions will have an impact on student learning via processes of teaching, learning, and collaboration. Researchers have to take these interrelationships into account when modelling system or school effects and drawing policy conclusions.

5.2 *Cross-cultural validity*

The aim of PISA is to compare student performance and its conditions across a large number of diverse countries. Cross-national studies do not only allow for analyses of policy approaches and other system-level inputs and processes and their relationships with performance, but also address issues relating to the transfer of findings from educational research to other cultures. Despite these benefits, cross-national studies also entail special methodological challenges. In addition to the regular tests of reliability and validity, cross-national research also needs to deal with the issue of equivalence and bias. Both refer to the question of time spent on tasks. Construct bias results from real differences in education systems, cultural norms, and interpretations of various phenomena, but may also be a consequence of translation errors or method bias. To address this potential problem PISA involves experts in all of the participating countries. Some of the actual differences in the meaning of the constructs are handled through national adaptations, but finally SEM analysis is needed to understand the level of cross-cultural construct equivalence, and thereby the kind of between-country comparisons – *i.e.* comparison of means, profiles, or correlations – that can be considered valid. For previous PISA cycles this issue has been addressed by using Multiple Group Confirmatory Factor Analysis (MGCFA) and examining whether the model has an acceptable fit across countries and whether the factor loadings are invariant. Results of these analyses show that PISA scales are quite equivalent across the whole sample of diverse cultures and education systems. For PISA 2012 the same analyses will be repeated, both for the existing and for any newly developed scales. The latter will only be kept for the Main Study if results are satisfactory.

Item bias

Item bias results from individual items that behave differently from other items measuring the same trait for particular groups. This can be detected using Item Response Theory (IRT) based Differential Item Functioning (DIF) methods. In previous PISA cycles, DIF analyses have been carried out both in the Field Trial and in the Main Study. A “dodgy item report” was sent to country representatives, and problematic items were revised (in the Field Trial) or – if necessary – excluded. PISA 2012 will use the same strategy.

Method bias

Method bias refers to a lack of comparability caused by the method used, *i.e.* by the sampling, by the administration or by the instrument itself. One instrument characteristic that may cause method bias is the use of Likert-type rating scales – a method that is frequently used in the PISA context questionnaires. Likert-type scales are especially susceptible to differences in response styles. Yet, culture is known to affect response styles. This interferes with the ability to examine mean scores for non-cognitive outcomes and to understand the relationships between contextual factors and performance (Hui & Triandis, 1985; van de Vijver, van Hemert, & Poortinga, 2008). Response styles may cause a lack of equivalence that can be detected with MGCFA or DIF analysis. However, as response styles often exert a similar influence on all items, additional analyses are required. A well known problem in PISA is the difference in the correlations of certain scales with performance at the between-student within-country and at the between-country levels (*e.g.* van de Gaer & Adams, 2010). For example, interest in mathematics correlates positively with mathematics performance within a country but the correlation at the between-country level is strongly negative. In other words, within countries higher-performing students report a higher level of interest in mathematics, as would be expected. At the between- country level, however, countries with higher mean performance have lower average levels of mathematics interest (*e.g.* Finland, Japan, and Korea) whereas higher levels of mathematics interest are recorded for countries with lower mean performance (*e.g.* Brazil, Indonesia, and Tunisia).

A more detailed analysis of method bias will help to preclude differences in response styles as a cause of differences in mean scores and correlations across countries. Three strategies will be used in PISA 2012 to address this problem. First of all, analysis will be carried out to examine whether statistical adjustments are feasible for correcting the problem for at least some of the scales. An example is the use of a mixture modelling approach (Rost, Carstensen & Von Davier, 1997) to cluster countries into categories that are similar with respect to response style. The second approach is to examine the feasibility of addressing the cross-cultural validity problem using a wide variety of new item types and other item manipulations, such as those suggested by Buckley (2009). These include reverse keyed items, forced-choice items, ranking items, items with various anchor labels so as to manipulate response scale usage, and manipulation of item position. Additionally, varying item formats will be used within a questionnaire (*e.g.* within a single questionnaire, to mix 3- vs. 4- vs. 5-point Likert-scale items). Finally, the effects of using vignettes or situational judgment items, and also alternative framings for questions (*e.g.* “compared to other students in your class,” vs. “in your country;” or, “compared to other fields of study”) can be analysed. Other design procedures are also suggested, such as identifying a multi-lingual sample within a country that could be used to centre scale points. All these methods require the collection of new data, and therefore could be accomplished during the Field Trial. To the extent that one or more design methods tried out during the Field Trial is successful in terms of reducing cross-cultural differences in response style effects, and increase cross-cultural measurement invariance, they will be employed in the Main Study in 2012.

5.3 Extending the sample to study change at the school level (international option)

The need for longitudinal data at the school level becomes clear in the light of reverse causality (also called reciprocal determinism, simultaneous effects, or recursive effects; Scheerens & Bosker, 1997). An example of reverse causality is when low performance causes certain policy decisions (*e.g.* remedial programs). In such a situation, cross-sectional data can produce negative correlations between performance and policy measures, even when controlling for student background.

Scheerens and Bosker (1997) argued that longitudinal research at the school level could provide an empirical basis for examining reverse causality effects, but they did not find any such studies. Improvements in the methodology of school effectiveness research (*e.g.* Creemers & Kyriakides, 2008), and the availability of fine-graded assessment and evaluation data for schools (including trend data over

several years) have greatly changed the state-of-the-art of this field of educational research. Researchers are aware of the phenomenon, and it is understood that for this reason, cross-sectional surveys are of limited value in explaining why some schools have more desirable outcomes than other schools.

One way of dealing with this issue within PISA is to allow for longitudinal analyses at the school level, by sampling the same schools twice in different PISA cycles in order to study school-level changes (*i.e.* changes in teacher qualifications, student socio-economic composition, student attitudes, and student performance – and how these are related over time).

5.4 Longitudinal extensions at the student level (recommended for future cycles)

Adding a longitudinal component to the PISA design setting would enable the examination of a number of interesting policy issues. Firstly it would allow for the estimation of value-added indicators, *i.e.* the effect of different variables on gains in student performance. While cross-sectional studies only examine relationships between variables, the direction of such effects can also be established with a longitudinal design. For example, the use of cross-lagged panel analysis or latent change models would allow for deciding whether self-efficacy has an effect on performance or whether performance has an effect on self-efficacy, or whether in fact reciprocal effects between both are found.

Some countries have already taken up the possibility of a longitudinal national extension of PISA, *i.e.* Australia (*Longitudinal Study of Australian Youth*) Canada (*Youth in Transition Survey*), the Czech Republic, Denmark, Germany, Switzerland (*Transitions from Education to Employment*) and Uruguay. Two approaches have been used:

- In Australia, Canada, Denmark, Switzerland, and Uruguay these studies have examined the transition from secondary to postsecondary education and/or to the labour market (*e.g.* Andersen, 2005; Bertschy, Boni & Meyer, 2008; Looker & Thiessen, 2008; Zoido & Gluszynski, 2009). Results contribute to the validation of the PISA performance tests by showing that both reading and mathematical literacy results predict indicators of educational pathways (drop-out, graduation, post-secondary education, university, unemployment, income, and so forth). This design further allows for the examination of effects of background and attitudinal variables on future life outcomes. For example, it has been shown that in Canada future educational pathways can be predicted by students' academic engagement, educational aspirations, extra-curricular activities, peer educational support, and parental support (Looker & Thiessen, 2008; Thiessen, 2007). In Australia, the individual perception of teacher-student relations (Curtis & McMillan, 2008) and teacher morale in school (Curtis & McMillan, 2008) are associated with later student outcomes. If more countries took up this option, it would be possible to examine whether the effects found in Canada and Australia can be generalised to other countries, and whether system-level characteristics (*e.g.* second chance programs) have an effect on students' educational paths as well.
- Another option is to examine the development of students' mathematics performance during secondary school, and the associations of school- and individual-level input and processes with latent change or growth in student competencies. Cross-sectional studies often find negative effects of different process variables, like homework support, remedial classes, or teacher cooperation. In part, this may be explained by the fact that these measures are often used in a remedial way. In other words, especially those students with learning difficulties are offered extra learning time. Likewise, in schools with many disciplinary problems, teachers may cooperate more closely because they feel more in need of support. Using a longitudinal design would allow to disentangle remedial use of measures from the actual effects of programs. More generally, it would improve the validity of conclusions drawn from analyses of the effects of policies and

processes on student performance, and thus provide a more valid knowledge base for policy decisions.

- Such a longitudinal design has already been realised by two German studies, the *Project for the Analysis of Performance Development in Mathematics* (PALMA) and *PISA International Plus* (PISA-I-Plus). These studies showed, for example, that the average student enjoyment in Grade 7 classrooms had a significant effect on teacher enjoyment in Grade 8, which, in turn, had an effect on student enjoyment in Grade 8, mediated by perceived teacher enthusiasm (Frenzel *et al.*, 2009). The study also revealed that classroom management had a significant effect on mathematics competency in Grade 10, and that mathematics performance in Grade 10 was significantly higher in schools with many active and/or discipline-oriented teachers as compared with schools with many passive teachers (Prenzel *et al.*, 2006).

Another example that has recently been discussed in the United States (Loveless, 2009), considers the issue of misplaced students, *i.e.* the negative effects of being enrolled in advanced courses on low-achieving students. If the critique put forward by Loveless (in the US context) holds, there should be a negative interaction effect of enrolment in special physics courses and certain risk factors (*e.g.* migration status, or low reading literacy) on mathematics performance. Within a cross-sectional design, this hypothesis cannot be tested properly. Longitudinal extensions would hence assist in answering complex policy questions.

5.5 **Final remark**

This framework for the context questionnaires in PISA 2012 has outlined how PISA can be developed further as a sustainable database for educational policy, and research. To this end, the framework started with a review of the general purpose and policy relevance of PISA and outlined the general knowledge base stemming from research in educational effectiveness. As PISA 2012 has again mathematics as its major domain, consideration has also been given to the specific issues involved in the contexts for teaching and learning mathematics. However, the framework's centrepiece is its aim to map out a design for the PISA context questionnaires that will be sustainable well into the future. To this end, the framework puts a system in place that accommodates recurring general material that is covered in every cycle and domain-specific material, which is covered every fourth cycle, thus allowing for trend analyses of general as well as domain-specific issues. In addition, the framework's system also allows for thematic extensions and specific foci to enable PISA to anticipate and incorporate new material or topics of interest to its audience. Based on the analyses of PISA data and other research findings, the framework proceeds to make specific recommendations regarding material to be included in PISA2012. The framework ends by making recommendation regarding aspects of design and analysis that are intended to build on the existing strengths of PISA and mapping an even stronger path into the future.

REFERENCES

- Abedi, J., M. Courtney, S. Leon, J. Kao, and T. Azzam (2006), *English Language Learners and Math Achievement: A Study of Opportunity to Learn and Language Accommodation (CSE Report 702, 2006)*, University of California, Center for the Study of Evaluation/National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, USA.
- Ajzen, I. (1991), "The Theory of Planned Behavior", *Organizational Behavior and Human Decision Processes*, Vol. 50, pp. 179-211.
- Alexander, K.L., D.R. Entwisle and L.S. Olson (2007), "Lasting consequences of the summer learning gap", *American Sociological Review*, Vol. 72, pp. 167-180.
- Amiot, C. and R.Y. Bourhis (2005), "Discrimination between dominant and subordinate groups: The positive-negative asymmetry effect and normative processes", *British Journal of Social Psychology*, Vol. 44, pp. 289-308.
- Andersen, D. (2005), "Four Years after Basic Compulsory Education 19-year-olds on Post Compulsory Education Choices and Paths", http://www.akf.dk/udgivelser_en/2005/4aarefter_grundskolen.
- Angelone, D., U. Moser, and E. Ramseier (2009), "Instruction Time and Performance. Analyses of the Importance of Instruction Time for Competencies in Science and Mathematics based on Swiss PISA 2006 Data", paper presented at the PISA Research Conference, Kiel, Germany, 14-16 September, http://www.pisaresconf09.org/user_uploads/files/context/room3/Angelone_Moser_Ramseier.pdf.
- Baker, D.P. (2009), "The Invisible Hand of World Education Culture", in G. Sykes, B. Schneider and D.N. Plank (eds.), *Handbook of Education Policy Research*, Routledge, New York, pp. 958-968.
- Baker, D.P., B. Goesling and G.K. LeTendre (2002), "Socioeconomic Status, School Quality, and National Economic Development: A Cross-National Analysis of the "Heyneman-Loxley effect" on Mathematics and Science Achievement", *Comparative Education Review*, Vol. 46, pp. 291-312.
- Baker, D.P. and G.K. LeTendre (2005), *National Differences, Global Similarities: World Culture and the Future of Schooling*, Stanford University Press, Stanford, California.
- Baker, M. L., J.N. Sigmon and M.E. Nugent (2001), "Truancy Reduction: Keeping Students in School", *Juvenile Justice Bulletin*, Office of Juvenile Justice and Delinquency Prevention, Washington, D.C..
- Bandura, A. (1977), *Social Learning Theory*, Prentice Hall, Englewood Cliffs, NJ, USA.
- Baumert, J., C.H. Carstensen and T. Siegle (2005), „Wirtschaftliche, Soziale und Kulturelle Lebensverhältnisse und Regionale Disparitäten des Kompetenzerwerbs“, in PISA-Konsortium Deutschland (ed.), *PISA 2003. Der zweite Vergleich der Länder Deutschland – Was wissen und können Jugendliche?*, Waxmann Verlag, Münster/New York/Berlin/ München.

- Baumert, J., W. Blum, M. Brunner, T. Dubberke, A. Jordan, U. Klusmann, S. Krauss, M. Kunter, K. Löwen, M. Neubrand and Y.-M. Tsai (2008), *Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz (COACTIV) - Dokumentation der Erhebungsinstrumente* (Baumert et al. (2009) Materialien aus der Bildungsforschung, No. 83), Max-Planck-Institut für Bildungsforschung, Berlin.
- Baumert, J., M. Kunter, W. Blum, M. Brunner, T. Voss, A. Jordan, U. Klusmann, S. Krauss, M. Neubrand and Y.-M. Tsai (2009), “Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress”, *American Educational Research Journal*, Vol. 47, pp. 133-180.
- Baye, A., C. Monseur and D. Lafontaine (2009), "[Institutional and Socioeconomic Factors Influencing the Reading Engagement](http://www.pisaresconf09.org/user_uploads/files/content/room1/Baye_Monseur_Lafontaine_2.pdf)", paper presented at the PISA Research Conference, Kiel, Germany, 14-16 September, http://www.pisaresconf09.org/user_uploads/files/content/room1/Baye_Monseur_Lafontaine_2.pdf.
- Belfield, C. and H. Levin (2002), “Education Privatization: Causes, Consequences, and Planning Implications”, *International Institute of Educational Planning*, UNESCO, Paris.
- Berezner, A. and P. Lietz (2009), “Implication for Preferred Model of Rotation on Cognitive Proficiency Estimate Generation Reporting and Subsequent Analyses”, QEG(0910)7.doc, paper presented at the QEG Meeting, Offenbach, Germany, 19-21 October.
- Berry, J. W. (1980), “Comparative Studies of Acculturative Stress”, *International Migration Review*, Vol. 21, pp. 491-511.
- Berry, J. W. (1990), “Psychology of Acculturation”, in R.W. Brislin (ed.), *Applied Cross-Cultural Psychology* (pp. 232-253), Newbury Park, CA, Sage.
- Bertschy, K., E. Boni and T. Meyer (2008), “Young People in Transition from Education to Labor Market. Results of the Swiss youth panel survey TREE”, update 2007, Basel, TREE, http://tree.unibas.ch/index.php?eID=tx_nawsecuredlandu=0andfile=fileadmin/tree/redaktion/docs/2007-12_Bertschy_Boeni_Meyer_TREE_Ergebnis-Update_engl.pdf&andhash=02fef28125ec37b71059c2179374b3ea.
- Blase, J. and J. Blase (1998), *Handbook of Instructional Leadership: How Really Good Principals Promote Teaching and Learning*, Corwin Press, Thousand Oaks, California.
- Blomhoj, M. and T.H. Jensen (2007), „What’s all the Fuss about Competencies? Experiences with Using a Competence Perspective on Mathematics Education to Develop the Teaching of Mathematical Modelling”, In Blum, W., P.L. Galbraith, H.-W. Henn, M. Niss (eds.), *Modelling and Applications in Mathematics Education*, 14 ICMI Study, Springer, New York, pp 45-56.
- Blum, W. and D. Leiss (2005), “How do Students and Teachers deal with Mathematical Modelling Problems? The Example “Sugarloaf””, in *ICTMA 12 Proceedings*.
- Blum, W. and D. Leiss (2007), “Investigating Quality Mathematics Teaching – the DISUM Project”, in C. Bergsten and B. Grevholm (eds.), *Developing and Researching Quality in Mathematics Teaching and Learning. Proceedings of MADIF 5*, SMDF, Linköping, pp. 3-16.
- Boaler, J. and J. Greeno (2000), “Identity, Agency and Knowing in Mathematical Worlds”, in J. Bowler (ed.), *Multiple perspectives on Mathematics Teaching and Learning*, CT Ablex Publishing, Westport.

- Borko, H., K.L. Kuffner, S.C. Arnold, L. Creighton, B.M. Stecher, F. Martinez, D. Barnes and M.L. Gilbert (2007), *Using Artifacts to Describe Instruction: Lessons Learned from Studying Reform-Oriented Instruction in Middle School Mathematics and Science*, CSE Technical Report 705, National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, USA.
- Bottani, N. and B. Favre (eds.) (2001), Open File: School Autonomy and Evaluation, *Prospects*, Vol. 31.
- Britner, S.L. and F. Pajares (2006), "Sources of Science Self-Efficacy Beliefs of Middle School Students", *Journal of Research in Science Teaching*, Vol. 43, pp. 485-499.
- Bromme, R. (1997), "Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers", in F. E. Weinert (ed), Reihe: Enzyklopädie der Psychologie, Titel: Pädagogische Psychologie, Bd. 3, *Psychologie des Unterrichts und der Schule* Hogrefe, Göttingen, pp. 177-212.
- Brookhart, S.M. (2009), *Exploring Formative Assessment*, ASCD Publishers, Alexandria, VA, USA.
- Brophy, J. (1992) (Ed.). *Planning and Managing Learning Tasks and Activities: Advances in Research on Teaching* (Vol. 3). Greenwich, CT: JAI Press.
- Brown, A.L. (1994), "The Advancement of Learning", *Educational Researcher*, Vol. 23, No. 8, pp. 4-12.
- Brunner, M., M. Kunter, S. Krauss, U. Klusmann, J. Baumert, W. Blum, M. Neubrand, T. Dubberke, A. Jordan, K. Löwen and Y.-M. Tsai (2006), „Die professionelle Kompetenz von Mathematiklehrkräften: Konzeptualisierung, Erfassung und Bedeutung für den Unterricht. Eine Zwischenbilanz des COACTIV-Projekts“, in M. Prenzel and L. Allolio-Näcke (eds.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms*, Waxmann Verlag, Münster, pp. 54-82.
- Buckley, J. (2009), „Cross-National Response Styles“, in International Educational Assessments: *Evidence from PISA 2006*, https://edsurveys.rti.org/PISA/documents/Buckley_PISAresponsestyle.pdf
- Carnoy, M., R. Elmore and L. Siskin (eds.) (2003), *The New Accountability: High Schools and High Stakes Testing*, Routledge Falmer, New York.
- Carroll, J.B. (1963), "A Model of School Learning", *Teachers College Record*, Vol. 64, pp. 723-733.
- Christenson, S.L. (2004), "The Family-School Partnership: An Opportunity to Promote the Learning Competence of all Students", *School Psychology Review*, Vol. 33, pp. 83-104
- Christiansen, B. and G. Walther (1986), "Task and Activity", in B. Christiansen, A. G. Howson and M.Otte (eds.), *Perspectives on Mathematics Education*, pp. 243-307.
- Chubb, J. and T. Moe (1990), *Politics, Markets, and American Schools*, The Brookings Institution, Washington, D.C., USA.
- Cleary, T.A. (1968), "Test Bias: Prediction of Grades of Negro and White Students in Integrated Colleges", *Journal of Educational Measurement*, Vol. 5, pp. 115-124.
- Coleman, J. (1988), "Social Capital in the Creation of Human Capital", *American Journal of Sociology*, Vol. 94, pp. 95-120.

- Common Core State Standards Initiative (2010), Common Core State Standards, <http://corestandards.org/>.
- Creemers, B.P.M. and L. Kyriakides (2008), *The Dynamics of Educational Effectiveness: A Contribution to Policy, Practice, and Theory in Contemporary Schools*, Routledge, London.
- Curtis, D.D. and J. McMillan (2008), *School Non-completers: Profiles and Initial Destinations. Longitudinal Surveys of Australian Youth*, Research Report 54, The Australian Council for Educational Research Ltd., Camberwell, Victoria, Australia.
- Desoete, A., and M. Veenman (Eds.). (2006). *Metacognition in mathematics education*. Hauppauge, NY: Nova Science.
- Flavell, J. H. (1979), "Metacognition and Cognitive Monitoring: A New Area of Cognitive-Developmental Inquiry", *American Psychologist*, Vol. 34, pp. 906-911.
- Frenzel, A. C., T. Goetz, O. Lüdtke, R. Pekrun and R. Sutton (2009), „Emotional Transmission in the Classroom: Exploring the Relationship between Teacher and Student Enjoyment”, *Journal of Educational Psychology*, Vol. 101, pp. 705–716.
- Fullan, M. (1992), *Successful School Improvement: The Implementation Perspective and Beyond*, Open University Press, Milton Keynes.
- Fullan M. and S. Stiegelbauer (1991), *The New Meaning of Educational Change*, 2nd ed., Teachers College Press, New York, USA.
- Gamoran, A. (1992), "The Variable Effects of High School Tracking", *American Sociological Review*, Vol. 57, pp. 812-828.
- Garofalo, J. and F.K. Lester (1985), "Metacognition, Cognitive Monitoring, and Mathematical Performance", *Journal for Research in Mathematics Education*, Vol. 16, pp. 163-176.
- Gustafsson, J.-E. (2007), "Understanding Causal Influences on Educational Achievement through Analysis of Differences over Time within Countries", in T. Loveless (ed.), *Lessons Learned: What International Assessments Tell Us about Math Achievement*, The Brookings Institution, Washington, D.C., USA, pp. 37-63.
- Hallinger, Ph. and R.H. Heck (1998). Exploring the principal's contribution to school effectiveness: 1980 – 1995. *School Effectiveness and School Improvement*, 9(2), 157-191.
- Hallinger, P. and J. Murphy (1986), "The Social Context of Effective Schools", *American Journal of Education*, Vol. 94 . No. 3, pp. 328-355.
- Hannaway, J. and M. Carnoy (eds.) (1993), *Decentralization and School Improvement: Can We Fulfill the Promise?*, Jossey-Bass, San Francisco, USA.
- Hanushek, E.A. and S.G. Rivkin (2007), "Teacher Quality", in E.A. Hanushek and F. Welch (eds.), *Handbook of the Economics of Education*, Elsevier, Amsterdam, pp. 1051-1078.
- Hanushek, E.A. and L. Woessmann (2009), "Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation", *NBER Working Paper No. 14633*, National Bureau of Economic Research, Cambridge, Massachusetts.
- Hanushek, E.A. and L. Woessmann (2010), "The Economics of International Differences in Educational Achievement", *IZA Discussion Paper No. 4925*, IZA, Bonn.

- Harris, A. and J.H. Chrispeels (eds.) (2006), *Improving Schools and Educational Systems: International Perspectives*, Routledge, London.
- Heck, R.H., and Ph. Hallinger (2005). The study of educational leadership and management. Where does the field stand today? *Educational Management Administration & Leadership* 33(2) 229-244.
- Hersbach, E. and P. Lietz (2010), "Towards a School Questionnaire in PISA 2012". QEG(1006)3.doc, paper presented at the QEG Meeting, Boston, Massachusetts, 21-23 June.
- Hesse, H. G., K. Göbel and J. Hartig (2008), „Sprachliche Kompetenzen von mehrsprachigen Jugendlichen und Jugendlichen nicht-deutscher Erstsprache“, in DESI-Konsortium (eds.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch*, Weinheim, Beltz, pp. 208-230.
- Hiebert, J., R. Gallimore, H. Garnier, K.B. Givven, H. Hollingsworth, J. Jacobs, A.M.-Y. Chui, A.M.-Y. Wearne, M. Smith, A. Manaster, E. Tseng, W. Etterbeek, C. Manaster, P. Gonzales and J.W. Stigler (2003), *Teaching Mathematics in Seven Countries: Results from the TIMSS 1999 Video Study*, US Department of Education, National Center for Education Statistics, Washington, D.C., USA.
- Hill, H.C., B. Rowan and D.L. Ball (2005), "Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement", *American Educational Research Journal*, Vol. 42, pp. 371-406.
- Hoover-Dempsey, K., and H.M. Sandler (1997), "Why do parents become involved in their children's education", *Review of Educational Research*, Vol. 67, pp. 3-42.
- Hopkins, D. (ed.) (2005), *The Practice and Theory of School Improvement: International Handbook of Educational Change*, Springer, Dordrecht.
- Hovey, J. D., and C.A. King,(1996), "Acculturative stress, depression, and suicidal ideation among immigrant and second-generation Latino adolescents", *Journal of the American Academy of Child and Adolescent Psychiatry* Vol. 35, pp. 1183-1192.
- Hui, C.H. and H.C. Triandis (1985), "Measurement in Cross-Cultural Psychology: A Review and Comparison of Strategies", *Journal of Cross-Cultural Psychology*, Vol. 16, No. 2, pp. 131-152.
- Husén, T. (1967), *International Study of Achievement in Mathematics*, Vol. 2, Wiley, New York, USA.
- Jehanghir, K. and C. Glas (2007), "The Predictive Power of the Variables in the Student Questionnaires. QEG(2007b)19", paper presented at the QEG Meeting, Arnhem, 8-9 June.
- Kao, G. and M. Tienda (1998), "Educational Aspirations of Minority Youth", *American Journal of Education*, Vol. 106, No. 3, pp. 349-384.
- Kaplan, D. (2009a), "Advances in Multilevel Latent Variable Models for PISA Data", paper presented at the PISA Research Conference, Kiel, 14-16 September, http://www.pisaresconf09.org/user_uploads/files/methodology/room1/Kaplan.pdf.
- Kaplan, D. (2009b), *Structural Equation Modeling: Foundations and Extensions*, 2nd Edition, Sage Publications, Newbury Park, CA, USA.
- Kaplan, D. (forthcoming), "Causal Inference in Non-Experimental Educational Policy Research", in D.N. Plank, W.E. Schmidt, and G. Sykes (eds.), in *AERA Handbook on Education Policy Research*, AERA, Washington, D.C., USA.

- Kaplan, D., J.-S. Kim and S.-Y. Kim (2009), "Multilevel Latent Variable Modeling: Current Research and Recent Developments", in R.E. Millsap (ed.), *The SAGE Handbook of Quantitative Methods in Psychology*, pp. 592-613.
- Kearney, C.A. (2008), "School Absenteeism and School Refusal Behavior in Youth: A Contemporary Review", *Clinical Psychology Review*, Vol. 28, pp. 451-471.
- Klieme, E., C. Pauli and K. Reusser (2009), „The Pythagoras Study: Investigating Effects of Teaching and Learning in Swiss and German Classrooms“, in T. Janik and T. Seidel (eds.), *The Power of Video Studies in Investigating Teaching and Learning in the Classroom*, Waxmann Verlag, Münster, pp. 137-160.
- Klieme, E., G. Schümer and S. Knoll (2001), "Mathematikunterricht in der Sekundarstufe I: Aufgabenkultur und Unterrichtsgestaltung [Mathematics instruction at lower secondary level. Task culture and instructional design]", in Bundesministerium für Bildung und Forschung (BMBF) (ed.), *TIMSS-Impulse für Schule und Unterricht. Forschungsbefunde, Reforminitiativen, Praxisberichte und Video-Dokumente*, pp. 43-57.
- Klieme, E. and K. Rakoczy (2003), „Unterrichtsqualität aus Schülerperspektive“, in Baumert, J. et al. (eds.), *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland*, Leske + Budrich, Opladen, pp. 333-359.
- Kobarg, M., T. Seidel, M. Prenzel, B. McCrae and M. Walker (2009), "Patterns of Science Teaching and Learning in an International Comparison", paper presented at the PISA Research Conference, Kiel, 14-16 September,
http://www.pisaresconf09.org/user_uploads/files/context/room3/Kobarg_etal.pdf.
- Koretz, D. (2008), *What Educational Testing Really Tells Us*, Harvard University Press, Cambridge, Massachusetts, USA.
- Krahn, H. and A. Taylor (2005), "Resilient teenagers: explaining the high educational aspirations of visible minority immigrant youth in Canada", *Journal of International Migration and Integration*, Vol. 6, No. 3/4, pp. 405-434.
- Krüger, M.L., B. Witziers, B. And P.J.C. Sleegers (2007). The impact of school leadership on school level factors: Validation of a causal model. *School Effectiveness and School Improvement*, 18(1), 1-20.
- Kunter, M. (2005). Multiple Ziele im Mathematikunterricht [Multiple goals in mathematics instruction]. Münster: Waxmann.
- Kunter, M., Y.-M. Tsai, U. Klusmann, M. Brunner, S. Krauss and J. Baumert (2008), „Students' and Mathematics Teachers' Perception of Teacher Enthusiasm and Instruction“, *Learning and Instruction*, Vol. 18, pp. 468-482.
- Kyriakides, L. and N. Tsangaridou (2004), "School Effectiveness and Teacher Effectiveness in Physical Education", Paper presented at the *85 Annual AERA Meeting*, American Educational Research Association, Chicago, USA.
- Lee, J. (2009), "Universals and Specifics of Math Self-Concept, Math Self-Efficacy, and Math Anxiety across 41 PISA 2003 Participating Countries", *Learning and Individual Differences*, Vol. 19, No. 3, pp. 355-365.

- Lee, V.E. and D.T. Burkham (2003). "Dropping out of High-School. The Role of School-Organization and Structure", *American Educational Research Journal*, Vol. 40, No. 2, pp. 353-393.
- Lee, V.E. and J.B. Smith (1997), "High School Size: Which Works Best and for Whom?", *Educational Evaluation and Policy Analysis*, Vol. 19, pp. 205-227.
- Leithwood, K., and C. Riehl (2005). What we know about successful school leadership. In W. Firestone & C. Riehl (Eds.), *A new agenda: directions for research on educational leadership* (pp.22-47). New York: Teachers College Press.
- Levin, H. (1978), "The Dilemma of Comprehensive Secondary School Reforms in Western Europe", *Comparative Education Review*, Vol. 22, pp. 434-451.
- Libman, Z. (2009), "Teacher Licensing Examinations – True Progress or Illusion?", *Studies in Educational Evaluation*, Vol. 35, No. 1, pp. 7-15.
- Lie, S. and M. Kjaernsli (2009), „Predictors for Students’ Choice of Science Careers”, paper presented at the PISA Research Conference, Kiel, Germany, 14-16 September, http://www.pisaresconf09.org/user_uploads/files/context/room3/Lie_Kjaernsli.pdf.
- Liebkind, K. (1996), "Acculturation and stress: Vietnamese refugees in Finland", *Journal of Cross-Cultural Psychology*, Vol. 27, pp. 161-180.
- Lipowsky, F., K. Rakoczy, C. Pauli, B. Drollinger-Vetter, E. Klieme and K. Reusser (2009), „Quality of Geometry Instruction and its short-term Impact on Students’ Understanding of the Pythagorean Theorem”, *Learning and Instruction*, Vol. 19, pp. 527-537.
- Lipnevich, A.A., C. MacCann, S. Krumm, J. Burrus and R.D. Roberts (forthcoming), "Math Attitudes in Belarusian and US Middle School Students", *Journal of Educational Psychology*.
- Looker, E.D. and V. Thiessen (2008), *The Second Chance System: Results from the three Cycles of the Youth in Transition Survey*, Learning Policy Directorate, Strategic Policy and Research, Quebec, http://www.hrsdc.gc.ca/eng/publications_resources/learning_policy/sp-836-04-08e/sp_836_04_08e.pdf.
- Loveless, T. (2009), *How Well are American Students Learning? With Sections on International Assessments, the Mismatched Math Student, and Urban Schools*, (The 2008 Brown Center Report on American Education), Brown Center on Education Policy at the Brookings Institute, Washington, D.C., www.brookings.edu/~media/Files/rc/reports/2009/0225_education_loveless/0225_education_loveless.pdf
- Lüttke, O., H.W. Marsh, A. Robitzsch, U. Trautwein, T. Asparouhov and B. Muthén (2008), "The Multilevel Latent Covariate Model: A New, more Reliable Approach to Group-Level Effects in Contextual Studies", *Psychological Methods*, Vol. 13, pp. 203-229.
- Ma, X. (1999), "Dropping out of advanced mathematics: The effects of parental involvement", *Teachers College Record*, Vol. 101, pp. 60–81.
- Ma, X. and N. Kishor (1997), „Assessing the Relationship between Attitude toward Mathematics and Achievement in Mathematics: A Meta-Analysis”, *Journal for Research in Mathematics Education*, Vol. 28, No. 1, pp. 26-47.

- Mau, W.-C. and L. Heim Bikos (2000), „Educational and Vocational Aspirations of Minority and Female Students: A Longitudinal Study”, *Journal of Counseling and Development*, Vol. 78, No. 2, pp. 186-194.
- McCluskey, C.P., T.S. Bynum and J.W. Patchin (2004), “Reducing Chronic Absenteeism: An Assessment of an Early Truancy Initiative”, *Crime and Delinquency*, Vol. 50, No. 2, p. 214.34
- McComb, E.M. and C. Scott-Little (2003), *A Review of Research on Participant Outcomes in After-School Programs: Implications for School Counselors*, ERIC Digest, Greensboro, NC: ERIC Clearinghouse on Counseling and Student Services, ED482765.
- Monk, D.H. (1994), “Subject Matter Preparation of Secondary Mathematics and Science Teachers and Student Achievement”, *Economics of Education Review*, Vol. 13, No. 2, pp. 125-145.
- Mullis, I.V.S. and M.O. Martin (2007), “TIMSS in Perspective: Lessons Learned from IEA’s Four Decades of International Mathematics Assessments”, in T. Loveless (ed.), *Lessons Learned. What International Assessments Tell Us about Math Achievement*, Brookings, Washington, pp. 9-36.
- Muthén, B. (2002), “Beyond SEM: General Latent Variable Modeling”, *Behaviormetrika*, Vol. 29, pp. 81-117.
- National Council of Teachers for Mathematics (NCTM) (1989), *Curriculum and Evaluation Standards for School Mathematics*, NCTM, Reston, Virginia, USA.
- Niss, M. (2003), “Mathematical Competencies and the Learning of Mathematics: The Danish KOM Project”, in A. Gagatsis and S. Papastavridis (eds.), *3rd Mediterranean Conference on Mathematical Education from 3-5 January 2003*, Athens: The Hellenic Mathematical Society, Athens, Hellas, pp. 115-124.
- Oakes, J. (2005), *Keeping Track: How Schools Structure Inequality* (2nd ed.), Yale University, New Haven, Connecticut, USA.
- Opdenakker, M.-C. and J. van Damme (2000), “Effects of Schools, Teaching Staff and Classes on Achievement and Well-Being in Secondary Education: Similarities and Differences Between School Outcomes”, *School Effectiveness and School Improvement*, Vol. 11, No. 2, pp. 165-196.
- Organisation for Economic Cooperation and Development (ed.) (2005a), *PISA 2003 Technical Report*, OECD, Paris.
- Organisation for Economic Cooperation and Development (ed.) (2005b), *School Factors Related to Quality and Equity. Results from PISA 2000*, OECD, Paris.
- Organisation for Economic Cooperation and Development (ed.) (2007a), *Education at a Glance*, OECD, Paris.
- Organisation for Economic Cooperation and Development (ed.) (2007b), *PISA 2006. Science Competencies for Tomorrow’s World*, OECD, Paris.
- Organisation for Economic Cooperation and Development (ed.) (2008), *Education at a Glance*, OECD, Paris.

- Organisation for Economic Cooperation and Development (ed.) (2009a), *Education at a Glance*, OECD, Paris.
- Organisation for Economic Cooperation and Development (ed.) (2009b), *Creating Effective Teaching and Learning Environments: First Results from TALIS*, OECD, Paris.
- Organisation for Economic Cooperation and Development (ed.) (2009c), *PISA 2009 Assessment Framework Key Competencies in Reading, Mathematics and Science*, OECD, Paris.
- [Pajares, F. and M.D. Miller \(1997\)](#), Mathematics Self-Efficacy and Mathematical Problem Solving: Implications of Using Different Forms of Assessment, *The Journal of Experimental Education*, Vol. 65, pp. 213–228.
- Perreault, S., and R.Y. Bourhis (1999), “Ethnocentrism, social identification and discrimination”, *Personality and Social Psychology Bulletin*, Vol. 25, pp. 92-103.
- Prenzel, M., J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, J. Rost and U. Schiefele (eds.) (2006), *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres*. Waxmann Verlag, Münster.
- Purves, A.C. (1987), “The Evolution of the IEA: A Memoir”, *Comparative Education Review*, Vol. 31, No. 1, pp. 10-28.
- Rakoczy, K., A. Buff and F. Lipowsky (2005), “[Teil 1: Befragungsinstrumente](#)“, in E. Klieme, et al. (eds.), *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie Unterrichtsqualität, Lernverhalten und mathematisches Verständnis*, GFPE, Frankfurt a.M.
- Rakoczy, K., E. Klieme and C. Pauli (2008). Die Bedeutung der wahrgenommenen Unterstützung motivationsrelevanter Bedürfnisse und des Alltagsbezugs im Mathematikunterricht für die selbstbestimmte Motivation [The impact of the perceived support of three basic psychological needs and of the perceived relevance of contents for students’ self-determined motivation in mathematics instruction]. *Zeitschrift für Pädagogische Psychologie*, 22(1), 25-35.
- Raudenbush, S.W. and A.S. Bryk (2002), *Hierarchical Linear Models: Applications and Data Analysis Methods*, Sage Publications.
- Ryan, R. M., and E.L. Deci (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, 25, 54-67.
- Rost, J., C.H. Carstensen and M. von Davier (1997), „Applying the Mixed Rasch Model to Personality Questionnaires“, in J. Rost and R. Langeheine (eds.), *Applications of latent trait and latent class models in the social sciences*, Waxmann, New York, pp. 324-332.
- Rowan B. E. Camburn, R. Correnti (2004), “Using teacher logs to measure the enacted curriculum in large-scale surveys: Insights from the Study of Instructional Improvement”, *Elementary School Journal*, Vol. 105, pp. 75–102.
- Rychen D.S. and L.H. Salganik (eds.) (2003), [Key Competencies for a Successful Life and a Well-Functioning Society](#), Hogrefe and Huber Publishers, Göttingen.

- Scheerens, J. (2000), "Improving School Effectiveness", *Fundamentals of Educational Planning Series*, IIEP, UNESCO, Paris, Vol. 68.
- Scheerens, J. and R.J. Bosker (1997), *The Foundations of Educational Effectiveness*, Pergamon, Oxford.
- Schmid, C.L. (2001), "Educational Achievement, Language-Minority Students, and the New Second Generation", *Sociology of Education (Extra Issue)*, pp. 71-87.
- Schmidt, W.H. and C.C. McKnight (1998), "What Can We Really Learn from TIMSS?", *Science*, Vol. 282, pp. 1831–1839.
- Schmidt, W.H., C.C. McKnight, R.T. Houang, H.C. Wang, D.E. Wiley, L.S. Cogan and R.G. Wolfe (2001), *Why Schools Matter: A Cross-National Comparison of Curriculum and Learning*, Jossey Bass, San Francisco, California, USA.
- Schmidt, W.H., R.T. Houang., (2007), "Lack of Focus in the Mathematics Curriculum: Symptom or Cause", in T. Loveless (ed.), *Lessons Learned: What International Assessments Tell Us about Math Achievement*, Brookings, Washington, USA, pp. 65-84.
- Schmidt, W.H. and A. Maier (2009), „Opportunity to Learn“, in G. Sykes, B. Schneider and D.N. Plank (eds.), *Handbook of Education Policy Research*, Routledge, New York, pp. 541-559.
- Schmidt, W.H. and C. McKnight (1995), „Surveying Educational Opportunity in Mathematics and Science: An International Perspective“, *Educational Evaluation and Policy Analysis*, Vol. 17, No. 3, pp. 337-353.
- Schneider, W. and C. Artelt (forthcoming), "Metacognition and Mathematics Education", in *ZDM/ The International Journal on Mathematics Education*.
- Schoenfeld, A. (2006), "Mathematics Teaching and Learning", in P.A. Alexander and I.H. Winne (eds.): *Second Handbook of Educational Psychology*, Routledge, London.
- Schoenfeld, A.H. (1992), "Learning to Think Mathematically: Problem Solving, Metacognition, and Sense-Making in Mathematics", in D. Grouws, (ed.), *Handbook for Research on Mathematics Teaching and Learning*. MacMillan, New York, pp. 334-370.
- Schoenfeld, A.H. (ed.) (1987), *Cognitive Science and Mathematics Education*, Erlbaum, Hillsdale, NJ, USA.
- Shulman, L. (1987), "Knowledge and Teaching: Foundations of the New Reform", in *Harvard Educational Review*, Vol. 57, No. 1, pp. 1-22.
- Slavin, R.E. (1990), "Achievement Effects of Ability Grouping in Secondary Schools: A Best-Evidence Synthesis", in *Review of Educational Research*, Vol. 60, No. 3, pp. 471-499.
- Snijders, T. and Bosker, R. (1999), *Multilevel Analysis*, Sage Publications, London, UK.
- Stanat, P. (2006), "Schulleistungen von Jugendlichen mit Migrationshintergrund. Die Rolle der Zusammensetzung der Schülerschaft", in J. Baumert, P. Stanat and R. Watermann (Hrsg.), *Herkunftsbedingte Disparitäten im Bildungswesen. Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit; vertiefende Analysen im Rahmen von PISA 2000*. Wiesbaden: VS Verl. für Sozialwiss, S. 189-219.

- Stanat, P. and G. Christensen, (2006), *Where Immigrant Students Succeed - A Comparative Review of Performance and Engagement in PISA 2003*, Organisation of Economic Co-Operation and Economic Development, Paris, Frankreich.
- Stein, M.K., R.A. Engle, M.S. Smith and E.K. Hughes (2008), “Orchestrating Productive Mathematical Discussions: Five Practices for helping Teachers move beyond show and tell”, in *Mathematical Thinking and Learning*, Vol. 10, p. 315.
- Stevens, F. (1993), “Applying an Opportunity-to-Learn Conceptual Framework to the Investigation of the Effects of Teaching Practices via Secondary Analyses of Multiple-Case-Study Summary Data”, in *Journal of Negro Education*, Vol. 62, No. 3, pp. 232-248.
- Stigler, J.W. and J. Hiebert (1999), *The Teaching Gap: Best Ideas from the World’s Teachers for Improving Education in the Classroom*. Free Press, New York, USA.
- Sui-Chu, H. and J.D. Willms (1996), “Effects of parental involvement on eighth-grade achievement”, *Sociology of Education*, Vol. 69, pp. 126–141.
- Teddlie, C. and D. Reynolds (eds.) (2000), *The International Handbook of School Effectiveness Research*, Routledge, New York, USA.
- Thiessen (2007), *The Impact of Factors on Trajectories that Lead to a High School Diploma and to Participation in Post Secondary Education*, Human Resources and Social Development Canada Publications Centre, Quebec, Kanada.
- Turner, S.L., J.C. Steward and R.T. Lapan (2004), “Family Factors Associated with Sixth Grade Adolescents’ Math and Science Career Interests”, *Career Development Quarterly*, Vol. 53, No. 1, pp. 41-52.
- Van de Gaer, E. and R. Adams (2010), “The Modeling of Response Style Bias: An Answer to the Attitude-Achievement Paradox?”, paper presented at the annual conference of the American Educational Research Association, Denver, Colorado, USA, 30 April – 4 May.
- Van de Vijver, F.J.R., D.A. Van Hemert and Y.H. Poortinga (eds.) (2008), *Multilevel Analysis of Individuals and Cultures*, Erlbaum, New York, USA.
- Van Landeghem, G., B. De Fraine and J. Van Damme (2005), “The Consequence of Ignoring a Level of Nesting in Multilevel Analysis: A Comment”, *Multivariate Behavioral Multivariate Behavioral Research*, Vol. 40, pp. 423-434.
- Vieluf, S., J. Lee and P. Kyllonen (2009a), “The Predictive Power of Variables from the PISA 2003 Student Questionnaire”, QEG(0910)5a.doc, paper presented at the QEG Meeting, Offenbach, Germany, 19-21 October.
- Vieluf, S., J. Lee and P. Kyllonen (2009b), “The Cross-Cultural Validity of Variables from the PISA 2003 Student Questionnaire”, QEG(0910)5b.doc, paper presented at the QEG Meeting, Offenbach, Germany, 19-21 October.
- Walberg, H.J. (1986), “Syntheses of Research on Teaching”, in M. C. Wittrock (ed.), in *Handbook of research on teaching* (3rd ed.), Macmillan, New York, USA, pp. 214–229.

- Wang, M.C., G.D. Haertel and H.D. Walberg (1993), "Toward a Knowledge Base for School Learning", in *Review of Educational Research*, Vol. 63, No. 3, pp. 249 -294.
- Wayne, A. and P. Youngs (2003), "Teacher Characteristics and Student Achievement Gain: A Review", in *Review of Educational Research*, Vol. 73, pp. 89-122.
- Wigfield, A., J.S. Eccles and D. Rodriguez (1998), "The Development of Children's Motivation in School Contexts", in A. Iran-Nejad and P.D. Pearson (eds.), in *Review of Research in Education*, American Educational Research Association, Washington D.C., USA, Vol. 23.
- Wilmers, N., D. Enzmann, D. Schaefer, K. Herbers, W. Greve and P. Wetzels (2002), *Jugendliche in Deutschland zur Jahrtausendwende: Gefährlich oder gefährdet?. Ergebnisse wiederholter, repräsentativer Dunkelfelduntersuchungen zu Gewalt und Kriminalität im Leben junger Menschen 1998 – 2000*, Nomos Verlagsgesellschaft, Baden-Baden, Germany.
- Witziers, B., R.J. Bosker and M.L. Krüger (2003). Educational leadership and student achievement: The elusive search for an association. *Educational Administrative Quarterly*, 39(3), 398-425.
- Woessmann, L. (2006), "The Complementarity of Central Exams and School Autonomy: Economic Theory and International Evidence", in E. Gori, D. Vidoni, E. Hanushek and C. Glenn (eds.), *Institutional Models in Education: Legal Framework and Methodological Aspects for a New Approach to the Problem of School Governance*, Wolf Legal Publishers Nijmegen, pp. 47-65.
- Yasushi, O. (2009), "Comparison of Attitudes toward Science between Grade 9 and 10 Japanese Students by Using the PISA Questions and Its Implications on Science Teaching in Japan", paper presented at the QEG Meeting, Offenbach, Germany, 19-21 October, http://www.pisaresconf09.org/user_uploads/files/content/room2/OGURA_Yasushi.pdf.
- Zimmer, K. (2008), "Contextual Information at Aggregate Levels: Potential Synergies with other OECD Data Collection Efforts", paper presented at the 25th Meeting of the PISA Governing Board, Warsaw, 6-9 April.
- Zoido, P. and T. Gluszynski (2009), "Pathways to Success. How Knowledge and Skills at Age 15 Shape Future Life Outcomes", paper presented at the PISA Research Conference, Kiel, Germany, 14-16 September, http://www.pisaresconf09.org/user_uploads/files/trends/room1/Gluszynski_Zoido.pdf.
- Zuzovsky, R. (2009), "Teachers' Qualifications and Their Impact on Student Achievement: Findings from TIMSS 2003 Data for Israel", in M. von Davier and D. Hastedt (Eds.), *IERI Monograph Series Issues and Methodologies in Large-Scale Assessments*, IER Institute, Hamburg / Princeton, Germany / USA, Vol. 2, pp. 37-62, [http://www.ierinstitute.org/IERI_Monograph_Volume_02\[1\].pdf](http://www.ierinstitute.org/IERI_Monograph_Volume_02[1].pdf).