PISA 2015

DRAFT QUESTIONNAIRE FRAMEWORK

Content

Introduction[1]

Providing indicators on effectiveness, equity, and efficiency of educational systems, setting benchmarks for international comparison, and monitoring trends over time are the most important goals of the Programme for International Student Assessment (PISA). In addition, PISA builds a sustainable database that allows researchers world-wide to study basic as well as policy-oriented questions on education, including its relation to society and economy.

In order to reach these goals, PISA not only needs reliable and valid measures for assessing cognitive student achievement (reading, mathematics, science literacy and other “life skills”), but also information on non-cognitive outcomes (e.g. students’ learning motivation), individual conditions (e.g. students’ cultural, ethnic, and socio-economic background), as well as structural and process characteristics of the institutional context (e.g. teaching practices and learning opportunities in classrooms, leadership and school policies for professional development, vertical and horizontal differentiation of the school system). It is this diverse set of constructs that is measured through questioning various stakeholders, namely students and school principals but (on an optional basis) also parents and teachers.

PISA 2015 will be the sixth wave of the programme. Since 2000, the so-called “Background Questionnaires” have gained substantially in importance. Meanwhile, questionnaires are of interest in their own right, beyond providing the “background” for reporting test results. PISA 2015 is a combination of student assessment in selected areas of achievement and context assessment on different policy-related issues. One reason for this change of perspective is that policy makers wanted the programme to report on a variety of issues that are relevant to professional practice, governance, and policy making in education. Thus, the scope of issues to be covered has gradually expanded since PISA 2000. In addition to student achievement and its relation with individual background, learning opportunities and non-cognitive outcomes, educational policies and practices are monitored through indicators. Also, data analysis and reporting have become more complex, allowing for more in-depth reporting. In addition to providing tables of indicators, patterns of input, process, and outcome variables are identified within and across countries; trends are reported, relations are explored, impact is estimated. These analyses require more sophisticated modelling approaches, and detailed data on contextual factors regarding students, schools, and educational systems.

Probably the most important characteristic of PISA, after 15 years, is the availability of trend data on the system level. PISA allows for the description of change in a country’s performance level over time, but also for the description of changes in non-cognitive outcomes, living conditions of adolescents and their families, professional practices and organisational structures for schooling. The more PISA moves towards repeated cycles of measurement, the more can be learned from examining the stability and variability of conditions, processes, outcomes and their relations: (a) Policy makers can use trend data for constant diagnosis and feedback. (b) The explanatory power of the study will increase because changes in performance can be interpreted and explained more substantively, taking changes in input and process into account (Gustafsson, 2008; Hanushek & Woessmann, 2010). (c) Last but not least, trend analyses are less likely to be biased by cultural issues. Quite often, policy makers and researchers have been reluctant to interpret “soft” constructs such as school climate, students’ motivation, teachers’ job satisfaction, or parents’ self-reported engagement, fearing their insufficient cross-cultural comparability. Now that trends are available, the focus is on change rates within countries rather than on cross-sectional comparisons of

[1] This document has been authored by Eckhard Klieme and Susanne Kuger at DIPF, based on discussions at the PISA 2015 Questionnaire Expert Group (QEG) and the PISA Governing Board (PGB). Chapters 2.2-2.5 summarize conceptual papers authored by members of the QEG and other experts contracted by Core 6 whose contributions are acknowledged in the respective chapter.

status. For example, the question whether school commitment is increasing or decreasing will be a relevant indicator within countries, and this indicator is not affected by response styles differing between countries. However, trend analyses require PISA to define a general set of constructs that will remain constant over several cycles in the future. This set of constructs will be referred to as "core content" in the following[2].

The present framework intends to explain the goals and the rationale for choosing appropriate questionnaire content, guiding both the questionnaire development and the forthcoming reports. The document is organized into two main parts: (1) defining the core content of the PISA questionnaires, and (2) elaborating a modular structure for broader coverage of policy issues.

The first part of this document links the current Framework to the overarching (cross-cycle) structure of PISA context assessment set out in the PISA 2012 Framework (OECD 2013, p. 168 ff.). The constructs that need to be covered for monitoring trends in education are revisited, with reference to the general background of Educational Effectiveness Research. Measures which have been used previously for initial reporting, for international indicators (published in "Education at a Glance"), and for secondary analyses are reviewed, culminating in an outline of core content that should be assessed in all cycles, for all participants.

The second and larger part of this document explores the full breadth of policy issues to be covered, structured by 19 modules, and explains how the most important modules—i.e. those judged as high priority modules by the PISA Governing Board—have been implemented in the Field Trial for PISA 2015. Here, detailed references to current research are provided.

The development of questionnaire content for PISA 2015 has been challenged not only by the intention of covering a broad array of policy issues, but also by the introduction of Computer-Based Assessment (CBA) as the preferred mode of administration for both cognitive tests and questionnaires. In addition to developing a large number of new questions, all content taken on from previous cycles has to be revisited, sometimes transformed into a CBA-friendly format (like using a "slider" that can be manipulated interactively instead of having the test taker fill in numeric information), and field trialled anew. For example, parents' occupational status, which is an important component of socio-economic background, used to be assessed by asking students to name their mother's and father's occupation; various alternatives for CBA equivalents will be tried out in the PISA 2015 Field Trial. Also, CBA allows for individual variation in the selection and ordering of questions. Using this feature, the PISA 2015 Field Trial will systematically compare variants of certain questions (including forced-choice formats and anchoring vignettes), study the impact of question order, item order, and the order of answering categories, and explore the reasons for non-response to certain questions. Finally, the analysis of process data based on logfile recording will deepen the understanding of response behaviour. These methodological studies will help improve the technical quality of the PISA data, and master the transition from paper-based assessment to CBA.

A large number of innovative questions related to all 19 modules will be tested in the Field Trial, providing a broad set of measures that can be used in the PISA 2015 Main Survey and/or in later cycles. Based on careful analysis of the Field Trial data and thoughtful discussion of priorities among experts and policy makers, constructs, questions, and items will be selected for inclusion in the PISA 2015 Main

[2] In the past, other, more technical notions of "core" have been used in the PISA Questionnaire Design. One approach used "core" to denote a set of variables in the Student Questionnaires which are measured for all students in a given PISA cycle – even in cases where different booklets are assigned to them. Another approach defined "core" as the set of variables used for imputing plausible values of test scores. A third approach referred to "core" as the set of domain-general variables, i.e. those not related to the major domain of assessment. Please note that in contrast to those definitions, the present Framework identifies "core content" as a set of conceptual constructs that defines the most basic context assessment necessary in PISA. This set of constructs (has been and) should therefore be included in all PISA cycles, albeit in some cases adapted to the major domain.

Survey. Severe cuts will be necessary: only about one third of the content field trialled for the Student Questionnaire and half of the content field trialled for the School and the Teacher Questionnaire can be kept. Once the content has been selected, this Framework will be revised and focused on the Main Survey content.

1. Defining the core of context assessment in PISA

Choosing from the many measures that might be incorporated in the PISA study design is a complex process, directed by the priorities that countries have set for the study, but also informed by educational research. One of the major forces that drive the PISA design in general is the rhythmic change of focus in the cognitive assessment: Reading literacy has been or will be the major domain of assessment in PISA 2000, 2009, and 2018, while mathematics is the focus of PISA 2003, 2012, and 2021; science takes the lead in PISA 2006, 2015, and 2024. Whatever serves as the major domain of cognitive assessment shall also be a major focus of "domain-specific" context assessment. However, there is a need for some stability in measurement to understand trends in education.

The Questionnaire Framework for PISA 2012 established an overarching framework that delineated core questionnaire content which should be kept comparable across cycles (OECD 2013, p. 189 ff.) to allow for continuous monitoring of educational systems. The overarching framework refers to domain-specific as well as domain-general measures assessing conditions, processes, and outcomes of education both for individual students and for schools. Finding an appropriate balance between these facets of the design is crucial for the long-term success of the PISA programme. In order to establish valid and reliable trends at the country level, it is important to implement a stable set of variables, which will be used as major reporting variables across PISA cycles[3].

This overarching framework is taken up in the following, along with specifying the constructs and measures in more detail and providing arguments that support the choice of core content for PISA 2015.

1.1. Outline of core content: constructs to be covered

Taking into account the goals of context assessment in PISA as stated in the introduction, strategic decisions made by the PISA Governing Board, the overarching framework developed for PISA 2012, and recommendations from the research literature, the current framework assumes that educational policy makers in participating countries need to be informed on four broad areas: outcomes, student background, teaching and learning processes, school policies and educational governance. In the following, these areas will be described on the level of conceptual constructs, while section 1.2 will refer to concrete measures that have been used in the past, and section 1.3 will specify a core set of measures for 2015 and future cycles. As stated above, these sections elaborate in more detail what has already been established in the Questionnaire Framework for PISA 2012.

1.1.1. Non-cognitive outcomes

The main challenge of PISA concerns measuring and documenting the outcomes of education that have been reached up to the age of 15 years. Educating a person basically means *fostering* his or her *individual development* as a unique, self-determined, knowledgeable person who gradually gains in ability to participate in society. As each PISA assessment is a cross-sectional study, PISA cannot capture developmental processes, but PISA serves as a snapshot of developmental status at the age of 15. This snapshot of course includes an assessment of literacy and life skills, but in addition to these cognitive outcomes, other factors are important outcomes, too. Success in school—and in life—depends on being

[3] From a technical point of view, it is also important to note that this stable set of background variables guarantees a strong set of conditioning variables used to impute measures of student proficiencies, as explained in the PISA Technical Reports (e.g. OECD 2005, p. 128pp.).

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 and Salganik, 2003). Educational research but also econometric analyses have shown that non-cognitive factors are most important for individual development as well as for success in life and well-being, and thus have an impact on individuals and society alike (Heckman, Stixrud and Urzua, 2006; Almlund, Duckworth, Heckman and Kauth, 2011).

Therefore, PISA addresses non-cognitive outcomes like attitudes, beliefs, motivation and aspirations, and learning-related behaviour, such as self-regulation, strategies and invested time. These non-cognitive outcomes are measured mainly within the Student Questionnaire (StQ), but also in the School Questionnaire (ScQ). They may be of a general nature, such as achievement motivation and well-being of students and drop-out rates of schools, or related to the domains of the cognitive assessment, such as reading engagement, interest in mathematics, or enjoyment of science. Domain-specific non-cognitive outcomes are also mentioned in the respective definitions of literacy, so this array of constructs serves as a link between test frameworks and context framework. Especially, students' self-efficacy beliefs—i.e. the strength of their belief in being able to solve tasks similar to the ones tested in the cognitive PISA tests—have been shown to be a strong correlate of student achievement both within and between countries.

1.1.2. Student background

In order to understand educational careers and to study equity issues within and across countries, family background variables such as the socio-economic status and ethnic background have to be taken into account. The distribution of educational opportunities and outcomes depending on these background variables shows whether countries succeed in providing equal opportunities.

PISA has become famous for its detailed, theory-based assessment of family background, socio-economic status, and immigration background. A lot of effort went into the definition and operationalization of individual student background indicators, finally leading to the establishment of a powerful, integrated indicator for students' economic, social, and cultural status (ESCS; Willms, 2006). The components of this indicator need to be assessed in as stable a way as possible across the PISA cycles. In addition, information on parental support helps understanding of how formal education and family background interact in promoting student learning.

Furthermore, PISA gathers retrospective and prospective information about educational pathways and careers across the lifespan. In recent years, researchers as well as public debate in many countries have stressed the importance of early childhood education (Blau and Curie, 2006; Cunha, Heckman, Lochner and Masterov, 2006). Therefore, PISA intends to catch at least some information on primary and pre-primary education.

On top of individual student background, the social, ethnic, and academic composition of the school he or she is attending has an impact on learning processes and outcomes. Therefore, PISA uses aggregated student data to characterize background factors on the school level in addition to structural factors such as school location, school type and school size.

1.1.3. Teaching and learning

School-based instruction is the core process of formal, systematic education. Therefore, policy makers need information on *teaching, learning, and organization of schools*. To increase the explanatory power of the study, assessment of teaching and learning will focus on the *major domain of assessment*, which in 2015 is science. The knowledge base of educational effectiveness research (Scheerens and Bosker, 1997; Creemers and Kyriakides, 2008) allows for the identification of core factors: teachers' qualifications, teaching practices and classroom climate, learning time and learning opportunities provided both within and out of school. For teaching processes, the focus should be on three basic dimensions (Klieme, Pauli and Reusser, 2009): (i) structure and classroom management, (ii) teacher support, (iii) cognitive challenge. Addressing teacher and teaching-related factors in PISA is a challenge, because sampling is by age rather than by grade or class. Nevertheless, aggregated student data and School Questionnaires can serve to describe the learning environment offered by schools.

1.1.4. School policies and governance

As policy makers have limited direct impact on teaching and learning processes, information on *school-level factors which help improve schools*, and thus indirectly improve student learning, shall have high priority. As with teacher and teaching variables, school effectiveness research has built a strong knowledge base showing that "essential supports" promote school effectiveness (Bryk, Sebring, Allensworth, Luppescu and Easton, 2010; see also Creemers and Reezigt, 1997; Scheerens and Bosker, 1997): professional capacity, with a focus on professional development; a well-organised curriculum; leadership and school management; parental involvement; school climate (truthful interactions between stakeholders, clear norms and shared values, high achievement expectations) and the use of assessment and evaluation for improvement. These factors shall be addressed within the PISA questionnaires as domain-general processes on the school level, assessed from different perspectives. In addition, school-level support for teaching the major domain shall be covered, such as the provision of laboratory space, information and communication technology (ICT) equipment and a coherent school curriculum for science education.

To meet policy requests directly, PISA also needs to address issues related to *governance* on the system level (Hanushek and Woessmann 2011; Woessmann, Lüdemann, Schütz and West, 2007). "Locus of decision making" measures and accountability practices describe main aspects of governance, namely the distribution of power and control between central and local stakeholders. Allocation, selection and choice as well as assessment and evaluation are the basic processes that policy makers and/or school administrators use to control school quality, to monitor and to foster school improvement. Some of this information can be gained from other sources (as documented in OECD's "Education at a Glance"), some can be assessed through the PISA School Questionnaire.

1.2. Previous use of PISA context data: measures that were important for analysis and reporting

In order to evaluate the importance of questionnaire content for PISA, it is worthwhile to look at previous cycles and how their data fed into analysis and reporting. Thus, the relevance of specific measures for policy making and research can be taken into account—in addition to the more abstract constructs mentioned before.

PISA data have been used for three types of analyses and reports: (1) Initial reports issued by the OECD (and, in some countries, by national centres) have been expanded from a single, mostly descriptive book reporting on PISA 2000 to the five-volume report on PISA 2009 published in 2010. In the following, this most recent and most sophisticated report will be used as a reference. (2) Indicators based on PISA data have been used for educational monitoring on a global as well as a national and regional scale. The most prominent set of indicators is OECD's annual report "Education at a Glance", which will be reviewed below. (3) Researchers'—including staff and fellows of OECD—involvement in publishing scientific research papers based on PISA data is continuously rising. As PISA 2015 will be focused on science, which also has been the major domain of assessment in PISA 2006, publications on science education using PISA 2006 data will be reviewed here.

Revisiting previous reports and publications will help identify those constructs and measures that have been used for analysis, while others have not been mentioned in any report so far. Obviously, this review can guide the selection of content for future cycles of PISA.

1.2.1. The PISA 2009 report

The PISA 2009 report used a lot of questionnaire material, especially for the diverse and rather sophisticated analyses presented in Volumes III and IV (see Table 1). In addition to student achievement, non-cognitive outcomes such as student engagement, cognitive and meta-cognitive strategies were studied in detail, and the impact of background variables, individual support factors, school level input, processes and policies, as well as system level factors was reported—all assessed in the Student and School Questionnaires.

Table 1: Measures based on questionnaires used in "PISA 2009 Results: What Students Know and Can Do" (OECD, 2010)

Volume I: Student Performance in Reading, Mathematics and Science
Student background: Gender
Volume II: Overcoming Social Background: Equity in Learning Opportunities and Outcome
Student background: ESCS, gender, immigration status, language spoken at home, age of arrival, country of origin
Individual support assessed through Parent Questionnaire: Parental support (at beginning of primary education/at age 15), pre-primary education (attendance, quality)
Volume III: Learning to Learn: Student Engagement, Strategies and Practice
Student background: ESCS, gender, immigration status, language spoken at home
Outcomes: Enjoyment of reading, time and material used for reading, metacognition (awareness of strategies), self-reported use of reading strategies (memorisation, elaboration, control)
Volume IV: What Makes a School Successful? Resources, Policies and Practices
Student background: SES, age of school entry, grade repetition
Student-reported processes: learning time (previous education, learning time at school, enrichment/remedial education, after-school lessons), teacher-student relations, disciplinary climate, teacher's stimulation of reading engagement
School input, policies and processes (reported by the principal): type of school (public/private), number of programmes, class size, educational resources (e.g. ICT, library), school responsibility for assessment and curriculum/for resource allocation, extra-curricular activities provided, school admittance/grouping/transfer policies, assessment practices/purposes, use of achievement data, school accountability, methods for monitoring teachers, teacher and student behaviour, parent involvement and expectations, leadership.

1.2.2. Education at a Glance

As documented in Appendix 1, PISA data, including questionnaire data, have been important sources for building and reporting educational indicators at OECD over the last decade. Numerous indicators based on PISA variables have also been formed in national reports.

Based on PISA questionnaires, (i) learning opportunities of 15-year old students, such as educational programmes and ICT equipment but also "soft" factors on school level such as school climate; (ii) non-cognitive outcomes such as students' attitudes and engagement; (iii) school and system level policies such as homework policies and use of school resources, as well as (iv) professional pedagogical practices like teacher support, achievement pressure, and preparation for science-related careers, have been documented. Most of the PISA-based indicators in OECD's yearly publication "Education at a Glance" can be found in chapter A, "The Output of educational institutions and the impact of learning", and chapter D, "The learning environment and organisation of schools".

1.2.3. Research publications based on PISA 2006

PISA 2006 was outstanding in its coverage of science-related learning opportunities, teaching practices, school-level context as well as student motivation, student engagement, beliefs and aspirations—scales covering a variety of domain-specific processes and outcomes. At the same time, analytical techniques and reporting were substantially improved. The initial report on PISA 2006, "Science Competencies for Tomorrow's World" (OECD, 2007), was a breakthrough in that it established multi-level modelling estimates of school effects explaining student outcomes and the degree of equity, both within and across countries.

Many additional examples of scientifically productive analyses can be found in the literature. For example, the ERIC international data base on educational research currently lists more than 100 peer-reviewed journal articles that are using PISA 2006 data to study science education. Sixteen out of these publications that used multivariate analyses are listed in Appendix 2. The majority of these papers discuss non-cognitive, domain-specific outcomes: re-scaling the questionnaire items, studying their structure within and across countries, analysing patterns of outcomes, looking at the impact of student and family background, identifying and explaining school effects.

1.3. Selecting and organising the core content

Addressing policy needs, and covering measures that have been used for reporting in previous cycles, a selection of core questionnaire content for PISA 2015 and beyond can be proposed. Table 2 organizes the suggested content according to the a model that has informed the design of international Large Scale Assessments for a long time (see, e.g. Purves, 1987, and OECD 2013, p. 173 ff.). The model allocates background, process, and outcome characteristics of education at respective levels of action (i.e., system level, school level embracing instruction/class/teacher factors, and individual student level).

Table 2: Measures to be included in the core context assessment for PISA
(Measures in italics must be adapted to the respective major domain, e.g. science in PISA 2015)

	Student and School Background	**Processes**	**Non-cognitive Outcomes**
System Level	()	**Governance:** Decision making, horizontal and vertical differentiation	(aggregated student data)
School Level	School location, type and size of school, amount and source of resources (incl. ICT) Social/ethnic/academic composition Class size, teacher qualification	**School policies:** Programmes offered, admission and grouping policies, allocated learning time, additional learning time and study support, *extra-curricular activities*, professional development, leadership, parental involvement, assessment/evaluation/account-ability policies, school climate (teacher and student behaviour) **Teaching and learning:** Disciplinary climate, teacher support, *cognitive challenge*	(aggregated student data) drop-out rate
Student Level	Gender, SES (parental education and occupation, home possessions, number of books at home), language and migration background, grade level, pre-primary education, age at school entry	Grade repetition, programme attended, learning time at school (mandatory lessons and additional instruction), *out-of school learning*	Domain-general non-cognitive outcomes (e.g. achievement motivation, well-being in school); Domain-specific non-cognitive outcomes (*motivation, domain-related beliefs and strategies* , *self-related beliefs*, *domain-related behaviour*)

The set of measures included in Table 2 comprises a core context design that (a) covers all construct areas mentioned in section 1.1, i.e. non-cognitive outcomes, student background, teaching and learning, school policies and governance; (b) allows for reporting all the analyses that have been included in initial reports, conducting all the research which is mentioned in Appendix 2, and calculating all indicators that have been developed for “Education at a Glance” (see section 1.2)[4]. It should be noted that Table 2 includes all questionnaire indices that have been shown to be strongly correlated with PISA achievement measures (e.g. number of books at home, SES, self-efficacy, and disciplinary climate), and thus will be instrumental in estimating test scores in PISA (“plausible values”). Therefore, this set of measures shall be considered for use in further PISA cycles including and beyond PISA 2015. Keeping this core design stable across cycles will enable trend analyses and complex modelling of system-level changes.

Most of the measures mentioned in Table 2 have already been used in previous cycles, mainly in PISA 2006 or PISA 2012, and thus they represent “trend” content that may be kept constant in the future. This

[4] With the exception of optional material such as the Parent Questionnaire, the ICT Familiarity Questionnaire, the Educational Career Questionnaire, and the PISA 2000 Questionnaire on cross-curricular competencies.

includes the science-specific measures which were mainly taken on from PISA 2006. When reading and mathematics were the major domain of assessment (PISA 2009 and 2012, respectively), different measures were used to represent the same overarching constructs:

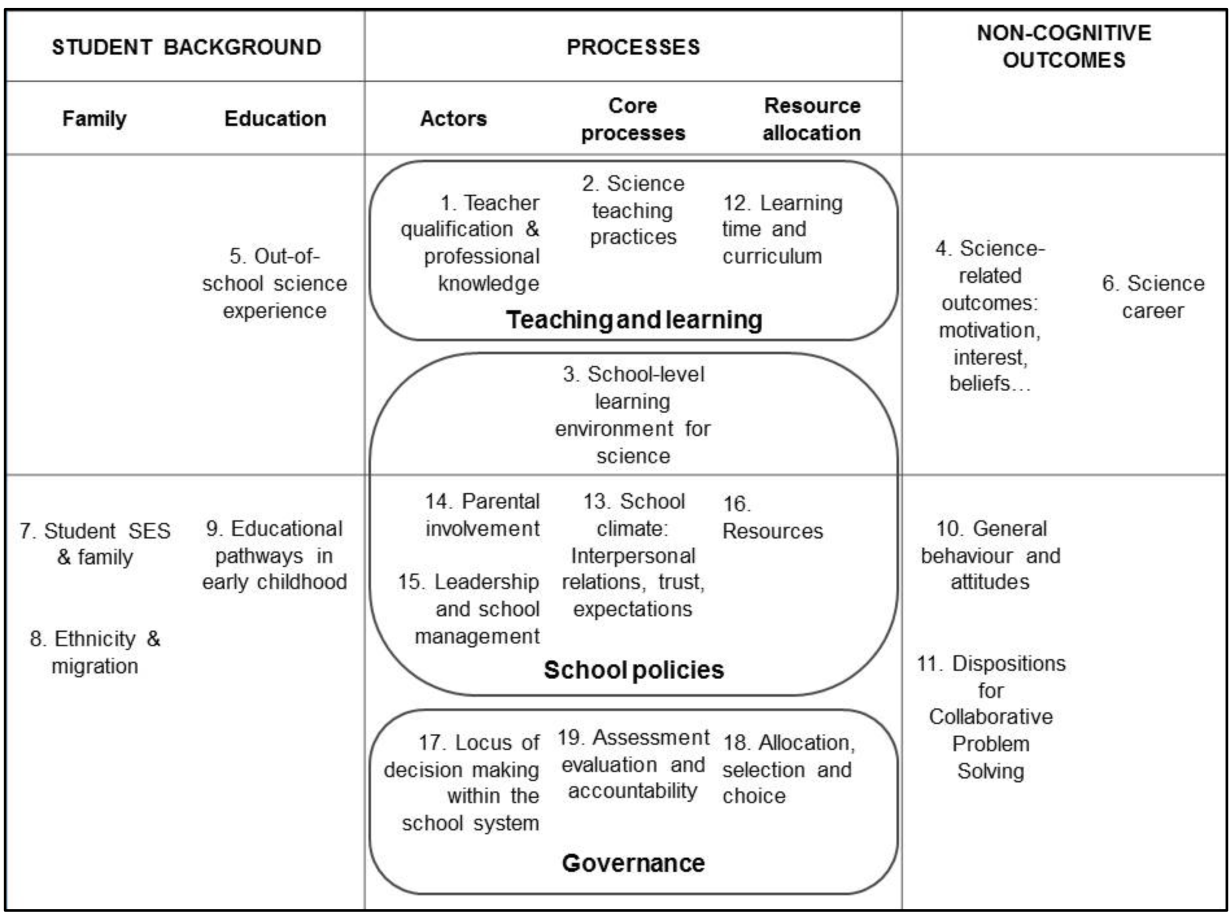
- *Cognitive challenge* in classrooms has been represented by teachers' stimulation of reading engagement (2009), Opportunities-to-learn (OTL)-Question types and experience with applied mathematical tasks (2012) and inquiry-based teaching and learning (2006, 2009).
- *Student motivation* has been operationalized by enjoyment of reading (2009), interest in mathematics (2012), and enjoyment of science (2006, 2015).
- *Domain-related behaviour* has been represented by reading for school and diversity of reading material (2009), mathematics work ethics and math behaviour (2012) and technology commitment (newly introduced in 2015).
- *Domain-related beliefs and strategies* have been represented by subjective norms on mathematics (2012) and value beliefs about science (2006, 2012); self-related beliefs have been represented by mathematics self-efficacy (2012) and science self-efficacy (2015); PISA 2009 introduced a measure of metacognition instead of reading-related beliefs.

2. Expanding the framework for broader coverage of policy issues

2.1. Modular approach to the PISA design

When the contractor for Questionnaire Development in PISA 2015 and the Questionnaire Expert Group started their work, they revisited the content areas described in section 1.1—non-cognitive outcomes, student background, teaching and learning, school policies and governance—and further differentiated them into 19 more fine-graded "modules", which finally were accepted by the PISA Governing Board (at its meeting in October, 2011) as the building blocks of the PISA 2015 design for context assessment. Figure 1 provides a schematic overview of this modular structure, positioning the modules within the overarching structure of background, process, and outcome characteristics.

Figure 1: Modular structure of the PISA 2015 context assessment design



Columns one and two summarize student background characteristics related to their family and the education they received, the three columns in the middle refer to educational processes on different levels (system governance, school policies, teaching and learning), and the columns on the right list various outcomes of education. In this figure, the lower part deals with domain-general topics, while the upper part includes modules that mainly deal with domain-specific (in this case: science-related) topics, including learning environment at the school level that specifically supports science education (Module 3) such as laboratories, science-related school curricula, collaboration among science staff, and the value attributed to science within the school community. Thus, the figure illustrates the combination of domain-general and domain-specific approaches to international large scale assessment which is typical for all PISA cycles, with either science, reading or mathematics being the major focus of assessment. As PISA integrates cross-curricular achievement measures like problem solving (in 2012) or collaborative problem solving (in 2015), appropriate non-cognitive outcomes are added (Module 11).

Traditionally, PISA treats the standard questionnaires (School Questionnaire and Student Questionnaire) separate from optional questionnaires that countries may choose to implement or not to implement. PISA 2015 will still keep them separate from an operational and reporting point of view, but the Questionnaire Expert Group intended to make the connections between standard questionnaires and options as transparent as possible. All modules are taken up in the standard questionnaires to some extent, while the options will be used to treat some modules in depth: Educational Career questions will address Modules 6, 9, and 12, while the ICT Familiarity Questionnaire contributes to Modules 7, 10, and 16, and the Parent Questionnaire provides content for Modules 9 and 14. The Teacher Questionnaire which has been added to the design for PISA 2015 will be relevant for Modules 1, 2, 3, 13, and 15. Thus, countries opting for any of these additional questionnaires will have additional information available for in-depth analysis of respective policy issues.

The expanded model will guide analysis and reporting in a systematic way:

- Each of the modules can be conceived as a thematic focus for analysis, as will be shown below in sections 2.2-2.5. Based on a comprehensive review of the corresponding research literature, each module shall cover main components that are relevant for a specific field of educational practice or policy making. Information gathered from students, school leaders, and – for countries choosing those options - parents and teachers will be combined to understand patterns and relations within countries and to compare between systems.
- Equity issues in education can be researched by studying outcomes in relation to background factors. (Un-)equal opportunities can be researched by studying schooling offered to various subgroups of students, while efficiency can be described as the relation between outcomes and resources.
- Models of educational effectiveness can be specified and tested by linking schooling to educational outcomes, controlling for background factors.

Every module represents a focus of policy making. Thus, the set of 19 modules covers a wide array of policy issues that are relevant across countries. This set is quite comprehensive, as can be seen by comparing the modular structure with literature on educational policy. For example, most topics treated by Sykes, Schneider and Plank (2009) in their state-of-the-art review of educational policy research are covered here.

To sum up, the modular approach to context assessment in PISA 2015 allows for a broad coverage of policy issues and related research questions. However, the PISA design sets strict limits to the length of questionnaires and in order to cover the requested breath of concepts, only some modules or constructs within modules can be focussed in more detail. To find relevant point of interest, PGB members were asked to indicate the top priority modules for further developmental work, based on policy relevance and the need for improvement from previous cycles. More emphasis will be devoted to those areas identified as priorities.

The areas receiving the highest votes for high policy relevance and need of further development work included non-cognitive outcomes (Modules 4 and 10), Teaching and learning (Modules 2, 12, and 1), and school policies (Modules 19 and 15). These modules will be discussed in detail in the following sections.

Considerable efforts have been made to include measures for those modules in the PISA 2015 Field Trial. Other modules will be discussed in a less detailed manner, as their content was taken over from previous cycles with little change.

2.2. Assessing non-cognitive outcomes[5]

This chapter summarizes the conceptual foundations for high priority modules 10 (Domain-general student behaviour and attitudes) and 4 (Science-related outcomes: Motivation, attitudes, beliefs, strategies) as well as those of lower priority modules 6 (Science career) and 11 (Dispositions for collaborative problem solving).

Traditionally, PISA assessed student outcomes in terms of achievement tests. Students' motivations, attitudes, beliefs, and behaviours were seen as important precursors of and predictors for scholastic performance, educational attainment, and labour market success. But educational policy and labour market policy are increasingly concerned about these "non-cognitive outcomes", because they are instrumental for personal growth, individual success, and society as a whole. Research findings have shown the predictive power of non-cognitive outcomes for success in secondary education, higher education and the workforce in general (e.g. Heckman, Stixrud and Urzua, 2006; Lindqvist and Vestman, 2011; Poropat, 2009; Richardson et al. 2012; Roberts et al. 2007). Also, professional and public debates often question the purely achievement-oriented approach that student assessments have mostly taken in the past. There is more to education than knowledge and cognitive skills. Therefore, non-cognitive outcomes become increasingly interesting as standalone outcomes in their own right. Non-cognitive dispositions are important goals, and they often function as moderators and mediators for relations of other constructs in the assessment. PISA offers a unique possibility of investigating complex relations between non-cognitive outcomes and achievement at the individual, school, and country level.

[5] This chapter is based on working papers submitted by Anja Schiepe-Tiska, Christine Sälzer, and Manfred Prenzel for Module 4, Jonas Bertling and Patrick Kyllonen for Module 10. Module 11 was developed in cooperation with Core 1 and the Collaborative Problem Solving Expert Group chaired by Art Graesser.

Table 3: Measures of non-cognitive outcomes included in the PISA 2015 Field Trial

Area	Science-related (Module 4)	Domain-general (Module 10)
Self	**Self-concept** **Self-efficacy** Dealing with uncertainty and ambiguity	Academic self-efficacy **(perceived control of success in school)** Test Anxiety Well-being: *Life satisfaction*, well-being at school (**sense of belonging**)
Interest, attitudes, and motivation	Interest in school subjects Interest in broad topics **Enjoyment of science** **Instrumental Motivation**	**Attitudes towards school: learning activities & outcomes** Achievement motivation
Beliefs and preferences	***Value of science: general, personal, labour market-related*** Occupational prestige Valuing scientific approaches to enquiry Epistemological beliefs ***Environmental Awareness / Optimism***	**Openness** **Perseverance** Industriousness Planning & Organisation Procrastination
Technology – ICT	Perceived competence Technology commitment Technology-related beliefs (benefits & harms)	Interest in ICT Perceived competence Autonomy in using ICT ICT in social interaction
Behaviour	Frequency of use of technical devices	Health School drop-out*

* Measured in the School Questionnaire. **bold** = trend measures. *Italics* = multiple measures in Field Trial

Previous PISA cycles have focused on domain-specific student attitudes and behaviours, for instance measuring attitudes towards reading and mathematics, mathematics self-concept, or maths anxiety; most of these scales display robust relations with student proficiency scores. This tradition is continued with Module 4 (Science-related outcomes) in PISA 2015. In addition, the current framework includes a larger set of domain-general non-cognitive student factors to broaden the coverage of relevant constructs, increase the policy relevance of the PISA 2015 database and acknowledge the increased interest in non-cognitive assessments both in policy and in research. Questions cover general attitudes towards school and motivation as well as more specific self-reported study behaviours and preferences. As in PISA 2012, anchoring vignettes (King and Wand, 2007) will be used for a better measurement to identify and correct for construct-unrelated response styles. This will increase the cross-cultural comparability of the derived indexes.

2.2.1. Science-related outcomes (Module 4)

As science is the major domain assessed in PISA 2015, students’ interest and motivation in Science, Technology, Engineering and Mathematics (STEM) subjects, plus related beliefs and behaviour are considered to be a very important assessment dimension. The second column in Table 3 provides an overview of the respective constructs in the broader domains of self, interest, attitudes, and motivation, beliefs and preferences, technology and ICT, and behaviour.

Motivation to learn science as well as value beliefs about science, beliefs about oneself as a science learner, and the use of technology are important policy and educational goals in many countries. Moreover, it has been shown that they are positively associated with students' performance in science (OECD, 2007). This relationship is reciprocal: Science-related beliefs and attitudes can be a consequence of higher science performance as well as causing higher science performance.

From a policy point of view, the shortage of skilled workers in technical and science professions—especially among females—has become a concern in recent years, and it is expected to rise in the future (European Commission, 2004, 2006; OECD, 2008). Therefore, PISA aims to identify how interested students are in science and how much they value science for themselves, the labour market, and the society. While some measures will allow for reporting trends compared to 2006, PISA 2015 adds a new focus on technology-related beliefs, attitudes and behaviour. This new focus is in line with the broader coverage of technology content in the PISA 2015 science assessment framework, and with policy concerns about human resources for technological development.

In addition, environmental issues are a global concern. Threats to the environment are prominently discussed in the media, and students are challenged to understand complex environmental issues. Further, students' levels of environmental awareness and optimism affect their engagement in environmental concerns and in turn affect the worlds' global climate, the economy, and the society as a whole. Therefore, PISA 2015 will take up some of the measures of environmental awareness that were developed for PISA 2006.

The following paragraphs provide relevant research background information and the different measures included in the PISA 2015 FT to cover these outcomes.

Self-related beliefs referring to science learning include three constructs: (1) science-related self-concept, (2) science-related self-efficacy, and (3) dealing with uncertainty and ambiguity. A positive self-concept and self-efficacy are highly related to motivation, learning behaviour, general expectations for the future, and students' performance (OECD, 2007). It has been found that self-concept is a significant predictor of undertaking science, technology, engineering, or math courses, while self-efficacy is a significant predictor of university entry in respective domains (Parker, Marsh, Ciarrochi, Marshall and Abduljabbar, 2013). Both constructs have been assessed in previous PISA cycles. Dealing with uncertainty and ambiguity involves students' attitudes towards unexpected or challenging events or things, which are common in scientific contexts. It also covers aspects of general curiosity. This construct is new to PISA 2015.

Motivation to learn science covers four constructs: (1) interest in broad science topics, (2) interest in school science topics, (3) enjoyment of science, and (4) instrumental motivation. Learning motivation that is based on interest and enjoyment is experienced as self-determinate and intrinsic (Krapp and Prenzel, 2011). It affects student engagement, learning activities, performance and career choices and it can be shaped by classroom instructions and parental motivation practices (Gottfried et al. 2009; Kunter, 2005; Rakoczy, Klieme and Pauli, 2008; Ryan and Deci, 2000). In addition, instrumental motivation is an important predictor for course selection, career choices, and performance (Eccles, 1994; Eccles and Wigfield, 1995; Wigfield, Eccles, and Rodriguez, 1998). All constructs have been used in previous PISA cycles, but the interest scale has been substantially revised and split into two measures.

Value beliefs about science include six constructs: (1) general value of science, (2) personal value of science, (3) the value of science on the labour market, (4) occupational prestige (5) epistemological beliefs, and (6) valuing scientific approaches to enquiry. *General value* of science implies the appreciation of the contribution of science to technology and social conditions of life. PISA 2006 showed that the majority of students valued science in general (OECD, 2007). *Personal value* of science is distinct from general values of science and students do not necessarily relate general values to their own life (OECD, 2007). However, higher levels of general and personal values are both related to better performance in science. *Value of science on the labour market* refers to students' perceptions as related to the importance of science for their future career. In PISA 2006, this question was only answered by the parents. In order to predict students' performance and career choices it may be more valid to ask students themselves. *Occupational prestige* assesses students' personal preferences or admiration for (academic) professions within or outside the field of science and technology, which is also relevant for students' career choices. *Epistemological beliefs* are closely related to students' general values of science and scientific inquiry (Fleener, 1996; Hofer and Pintrich, 2002). They include beliefs about science as an evolving and changing subject and how individuals justify knowledge (Conley, Pintrich, Vekiri, and Harrison, 2004). So far in PISA, epistemological beliefs have been assessed in the domain of mathematics but not in the domain of science. *Scientific approaches to enquiry* have been highly successful in generating new knowledge. Moreover, the core value of scientific enquiry and the enlightenment is the belief in empirical evidence as the basis of rational belief. They have not been assessed in previous PISA cycles.

Environmental issues are a distinct area of beliefs that is covered by two constructs: (1) environmental awareness, and (2) environmental optimism. PISA 2006 showed that students from more advantaged socio-economic backgrounds reported higher levels *of awareness of environmental issues* and that they are linked with students' science performance (OECD, 2007). In addition, students reported low levels of *environmental optimism*, which was negatively associated with students' performance.

Technology plays an important role in the PISA 2015 Assessment Framework for Science, in addition to more traditional areas such as physics, chemistry, or biology. Accordingly, the Science Expert Group and the Questionnaire Expert Group were eager to develop new measures for student beliefs and attitudes related to technology: (1) perceived competence as related to technical devices, (2) technological commitment, (3) weighting benefits and harms of technologies, and (4) everyday use of technical devices. *Perceived competence* asks about how confident students feel when they use new technical devices. In addition, *technology commitment* assesses students' acceptance of new technologies and their locus of control as related to new technologies. Studies have found that men show higher levels of technological commitment than women (cf. Neyer, Felber, and Gebhardt, 2012). *Weighting benefits and harms of technologies* informs about students' attitudes as related to new technology, i.e. whether they tend to perceive the benefits or the harms of new technologies. Finally, *use of technical devices* is a behaviour measure of the role technologies are playing in students' everyday life. All of these constructs are new to PISA 2015.

2.2.2. Domain-general student attitudes and behaviours (Module 10)

Domain-general attitudes, beliefs and behaviours can be defined as a set of student factors or constructs that cut across curricular topics, or that are independent of curricula, but that are nevertheless important for and reflective of success in education. PISA 2015 does not intend to measure global dispositional traits but behavioural tendencies and preferences that are relevant in the context of learning and that can be conceptualized as outcomes of education in general, and schooling in specific.

As shown in Table 3, right column, the content of Module 10 can be grouped into broader areas which basically parallel the science-related outcomes, widening the goal of assessment well beyond science and technology. The following paragraphs provide links to previous work at OECD and other research, especially focusing constructs that are new to PISA such as well-being, health, and time use.

Self-related beliefs and attitudes towards school: Generalized beliefs about one's own success or failure in academic learning have been shown to be strong predictors for further effort and success, including test scores in student assessments. PISA 2015 takes up two such measures: A scale named

“Perceived Control of Success in School” introduced in PISA 2012, which directly measures academic self-efficacy beliefs, as well as a revised (and generalized) version of a (test) anxiety scale that was predictive for mathematics achievement in previous PISA cycles. Furthermore, a new indicator for generalized achievement motivation (e.g. “I want to be the best, whatever I do”) has been introduced, and two sets of items from PISA 2012 have been taken up, asking the student to evaluate his or her previous experience in school (e.g. “School has taught me things which could be useful in a job”) and learning activities in general (e.g. "Trying hard at school is important”). Altogether, these self-related and attitudinal measures will help to understand motivational processes in school from a general, domain-independent perspective. Anchoring vignettes (King and Wand, 2007) are implemented to control for response biases and increase cross-cultural equivalence. Depending on the results of the Field Trial, some of these constructs shall be chosen for continuous monitoring of school-related motivations in further cycles.

Subjective well-being: Subjective well-being can be defined as “Good mental states, including all of the various evaluations, positive and negative, that people make of their lives and the affective reactions of people to their experiences” (OECD, 2013, p. 10). This definition encompasses three elements of subjective well-being: *life evaluation*—one’s reflective assessment of one’s life (including the single “general life satisfaction” question); *affect*—an emotional state, typically at a particular point of time; and *eudaemonia*—a sense of meaning and purpose in life. The growing recent interest from researchers and policy makers in this construct has resulted in recommendations to statistical agencies to “incorporate questions on subjective well-being in their standard surveys to capture people’s life evaluations, hedonic experiences and life priorities” (Stiglitz, et al. 2009, p. 216). OECD (2013) has responded to this charge in providing guidelines on measuring subjective well-being. To date, 27 out of 34 OECD national statistical offices have committed to collecting at least the minimal information proposed by the OECD guidelines (the single “general life satisfaction” question). For PISA 2015 in keeping with the guidelines information on all three elements, *life evaluation*, *affect*, and *eudaemonia*, shall be collected. The guidelines also suggest that it is perfectly appropriate to collect such information from 15 year olds, and even younger students because the evidence suggests that they are “capable of responding effectively to subjective well-being questions from as young as age 11 with respect to measures of life evaluation and affective state” (p. 152).

Preferences in learning and working: Students’ *openness* for problem solving (e.g. enjoy solving complex problems, seek explanations for things), their *planning and organization* behaviours (e.g., keep to-do lists, keep notes for subjects; finish assignments on time, not to do things at the last minute), persistence even on difficult tasks (*perseverance*, e.g. not to put off difficult problems, not to give up easily), general work ethics (*industriousness*, e.g. prepare for class, work consistently throughout the school year) and low level of *procrastination* are not only among the strongest non-cognitive predictors of school grades (see recent meta-analysis by Richardson et al. 2012), but also important predictors of success in higher education and the workforce in general (e.g. Heckman, Stixrud and Urzua, 2006; Lindqvist and Vestman, 2011; Poropat, 2009; Roberts et al. 2007). Two of these constructs were already successfully implemented in PISA 2012 (openness and perseverance).

ICT-Technology: Whereas Module 4 (described above) deals with beliefs and attitudes directly connected to the educational domain of science and technology, Module 10 covers one aspect of technology which is relevant in everyday live and across all educational domains, namely ICT technology. ICT-related behavioural characteristics and motivational attributes can be regarded as domain-general student outcomes. Since ICT subsumes a broad range of devices, it may play a role across all educational domains. Following the OECD’s DeSeCo project and the 21st Century Skills Initiative, students should exhibit general skills related to information, media and technology above and beyond the traditional core subjects (OECD, 2005b; Partnership for 21st Century Skills, 2008). PISA 2015 will assess students’ *interest in ICT*, (self-determined) practices of *ICT use*, their *perceived competence and* *autonomy* in using ICT, and a specific question on *use of social media* within the ICT Familiarity Questionnaire.

Health: This area addresses health literacy and the practice of healthy behaviours, particularly knowledge of the benefits of, attitudes towards, and behaviour pertaining to a good nutritious diet, regular exercise, and health-related lifestyle choices. Research has shown that health literacy affects students' decisions related to diet, exercise, and lifestyle, and that those decisions in turn are related to academic performance. Poverty and low socioeconomic status is associated with poor health outcomes (Spurrier, Sawyer, Clark, and Baghurst, 2003). Research has shown that physical activity interventions can improve academic performance due to the activity itself and to the displacement of physical activity for less healthy more passive activities (Salmon, Ball, Crawford, Booth, Telford, and Hume, 2005). Nutrition and physical exercise are health factors that can be (positively as well as negatively) influenced by teacher behaviours and school practices. For example, school meal provision programmes may contribute to a more positive experience of school, improve concentration and reduce illness (Kristjansson, Petticrew, MacDonald, Krasevec, Janzen, Greenhalgh, et al. 2009). The approach taken here to measure health factors in PISA 2015 is adapted from and closely aligned with the international health or social surveys: *Global school-based student health survey* (GSHS; WHO), National Youth Physical Activity and Nutrition Study (NYPANS), Strengths and Difficulties Questionnaires (SDQ, e.g. Goodman, 1997), STEPS (see e.g. Bonita et al. 2001), and World Health Survey (WHO). In particular, the GSHS is administered to an age group similar to the PISA population. These measures are reflective of cultural norms to some degree, of course, but no more so than are other measures collected on the questionnaire, including for example, truancy, number of books in the home, parental education, teacher support, self-confidence, locus of control, or perseverance.

2.2.3. Non-cognitive outcomes related to Collaborative Problem Solving (Module 11) and science career (Module 6)

The PISA 2015 Student Questionnaire takes up a number of questions about career aspirations, both general and science-related, that have been used in 2006.

In order to cover dispositions related to the new domain of assessment introduced in PISA 2015, namely Collaborative Problem Solving, a set of items on valuing team work, cooperating, guiding others, and negotiating has been developed based on research by Wang and colleagues (2009). Behavioural measures addressing student experience with collaboration-type activities inside and out of school, both technology-based and other, have been added, while the Teacher Questionnaire covers types of activities and grouping, teacher support and rewards from yet another perspective.

2.3. Assessing teaching and learning processes[6]

This chapter summarizes the conceptual foundations for high priority modules 2 (Science teaching practices), 12 (Learning time and curriculum) and 1 (Teacher qualifications and professional knowledge) as well as those of lower priority module 5 (Out-of-school science experience).

Teaching and learning are at the heart of schooling. Most cognitive and non-cognitive, curricular and cross-curricular goals of school education are achieved—or impeded—by the way students and teachers

[6] This chapter is based on working papers submitted by Katharina Müller, Manfred Prenzel, and Tina Seidel for Module 2, Susanne Kuger for Module 12, Eckhard Klieme, Franz Klingebiel and Svenja Vieluf for Module 1.

interact in classrooms. While teaching is the core process in schools, the curriculum determines its content, and professional teachers are the force who implement the curriculum, orchestrate learning activities, and thus arrange for quality learning time.

PISA has been designed as a yield-study, assessing life-skills and broad areas of literacy rather than curricular domains, sampling a birth cohort rather than a grade level or intact classrooms. Thus, it might be questioned why this programme should address teaching and learning processes at all. However, there is ample evidence that teaching and learning activities are the best predictor of student competencies, whatever their character might be. So, if PISA is to inform educational policy making at the system and the school level, it must cover this important area. Clearly, the PISA study should focus on more general and internationally comparable constructs, rather than fine-grained content matter. Therefore, Module 2 describes science education by broad lists of teaching and learning activities, including—but not limited to—inquiry-based teaching and use of technology. In addition, general dimensions of teaching quality such as well-structured instruction, classroom management, support, and cognitive activation are applied to science education. Furthermore, Module 12 covers learning time—including non-mandatory, additional instruction within and out-of school—as well as coherence, focus, and rigour of the science curriculum. Finally, the teaching force will be described in terms of initial education, beliefs, and professional development (Module 1).

2.3.1. Science teaching practices (Module 2)

According to the PISA approach to scientific literacy, the main task of science teaching is to foster students' capacity to explain phenomena scientifically, to understand scientific enquiry and to interpret scientific evidence. The key topic of the framework outlined in the following is the question to what extent schools are mastering this task.

A number of processes at the classroom level have been found to be relevant for effectiveness in science education. In this framework both domain-specific instructional approaches and activities, and more general dimensions of instructional quality are combined, as they are equally suited to support learning activities and to describe processes on the classroom level. However, in PISA 2015, all questions about teaching and learning activities are framed within the context of school science, sometimes even referring to one specific course. The aim is to describe science teaching in the classroom by country-specific profiles of teaching practices and to investigate their relation to students' outcomes.

Analyses based on PISA 2006 show, that a student's outcome can be predicted by different profiles of practices in teaching (Kobarg et al. 2011). While some teaching patterns are related to high performance, others are related to high student interest and motivation. The results indicate that the items and scales for science teaching practices are applicable to in-depth descriptions of science teaching in the classroom. Moreover, a comparison of the patterns allows detailed analyses concerning both students' science performance and students' interest in science topics for all countries (Kobarg et al. 2011; Prenzel, Seidel, and Kobarg, 2012). The teaching practices items are developed and chosen in order to discriminate between different patterns of teaching. An overview of dimensions and constructs of Module 2 is shown in Table 4.

Table 4: Science teaching practices

Domain-specific instructional approaches and activities	**General dimensions of teaching quality**
Inquiry-based teaching practices • epistemic activities • social activities • procedural activities • real life applications	Teacher's instruction Disciplinary climate
Teaching and learning activities (assessed in Student Questionnaires and – for countries choosing this option – in the Teacher Questionnaires) • hands on activities • reading science • talking about science • writing science	Cognitive activation Teacher support Perceived support of competence
Use of ICT in science lessons	Perceived support of autonomy

Teaching and learning activities: Research has shown that *inquiry-based teaching practices*, which play a significant role in science education, have a positive effect on student learning. Especially students' engagement in the cognitive dimensions of inquiry and teacher-led inquiry activities are of importance (Furtak, Seidel, Iverson and Briggs, 2012). Furthermore, inquiry-based instruction seems not only to improve achievement (Blanchard et al. 2010) but also attitudes towards the subject and transferable critical thinking skills (Hattie, 2009). A renewed interest in embedding science teaching and learning in contexts that are real and meaningful for learners can be observed (Fensham, 2009; King and Stephen, 2012). Scientific argumentation as a central goal of science education (Osborne, 2012) needs classroom situations with sufficient opportunities for social interaction. Particularly instruction that emphasises students' active thinking and drawing conclusions from data seems to be beneficial for students' development (Minner, Levy and Century, 2010). According to these findings and the analysis of the PISA 2006 items (Kobarg et al. 2011, Taylor, Stuhlsatz, and Bybee, 2009), an improved framework for inquiry-based teaching is introduced distinguishing the cognitive and social activities of the student, and the guidance provided to students by their teacher (Furtak, Seidel, Iverson, and Briggs, 2012). The dimension *inquiry-based teaching practices* covers four constructs: epistemic activities, social activities, procedural activities and real life applications. Focusing on both student-led and teacher-led activities that support learning-processes in science education, the original item pool from PISA 2006 was reduced and adapted, and in a second step augmented with some new items to depict the four constructs. The development of new items particularly seized upon the role of argument in school science (Jiménez-Aleixandre and Erduran, 2007, Osborne, 2012). Items with a positive relation to performance (Taylor, Stuhlsatz, and Bybee, 2009) were kept.

In addition to the inquiry-based teaching practices, *Teaching and learning activities in science lessons* and *use of ICT in science lessons* focus on executive activities and teaching methods in science lessons and broaden the perspective on domain-specific practices. The purpose is to obtain student-reported information about their actions in school science lessons and to get a realistic picture of what is going on in science classrooms—including classrooms with little inquiry based learning. The dimension *teaching and learning activities in science lessons* covers four constructs: hands on activities, reading science, talking science, and writing science. This scale as well as the *use of ICT in science lessons* scale has been newly developed for PISA 2015 (based on Ainley, Eveleigh, Freeman, and O'Malley, 2010) and is meant to address the use of digital devices in science lessons.

The student perspective on science teaching is complemented by the Teacher Questionnaire, for those countries which participate in this option: Science teachers are asked to describe their teaching practices by a parallel version of *Teaching and learning activities in science lessons*. Both perspectives may be combined and compared on the school level.

Dimensions of teaching quality: Numerous studies confirmed the impact of three basic dimensions of instructional quality on students' cognitive and motivational development: clear, well-structured classroom management, supportive, student-oriented classroom climate, and cognitive activation with challenging content (Klieme, Pauli, and Reusser, 2009). These dimensions of instructional quality were covered by the dimensions *teacher's instruction*, *disciplinary climate*, *cognitive activation*, and *teacher support*. All constructs have been used in previous PISA cycles. The *teacher's instruction* scale asks for student-reported information about direct teacher-led instructions focusing on goal orientation and different forms of instructional scaffolds. As a school climate variable, the purpose of the *disciplinary climate* question is to gain information on the structure and efficiency of classroom management, which can be seen as a prerequisite for student learning. The purpose of the *cognitive activation* question is to obtain information about thought-provoking mathematics instructions. Assuming that the level of cognitive challenge is determined by the type of problem and the way of its implementation in the lesson, the concept of cognitively activating tasks was slightly adapted for science teaching. Teacher support measures how often the teacher helps students with their learning (OECD, 2004). Research has shown, that the scale is positively related to students' interest (Vieluf, Lee and Kyllonen, 2009)

Two subdimensions were newly developed in addition to the teacher support scale. The constructs *perceived support of competence*, and *perceived support of autonomy* refer to the Self-determination theory of motivation (Deci and Ryan 1985, 1993; Ryan and Deci, 2002) and ask about the teachers' support concerning basic psychological needs (Prenzel, Kristen, Dengler, Ettle, and Beer, 1996). The teacher supports students' competence by providing them opportunities to regulate learning processes on their own. Autonomy is supported when students can determine and realize their personal goals and interests. Several studies show positive consequences for students' cognitive and emotional processes (Ryan and Deci, 2000; Kunter, 2005)

2.3.2. Learning time and curriculum (Module 12), including out-of-school science experience (Module 5)

The learning time and curriculum, which students experience in their course of educational experiences, are closely related to student outcome (e.g. Schmidt and Maier, 2009; Abedi et al. 2006; Scherff and Piazza, 2008).

Learning time has proven to be a central factor in student learning and achievement (Gándara et al. 2003; Patall, Cooper and Allen, 2010; Scheerens and Bosker, 1997; Seidel and Shavelson, 2007). Such positive relations were replicated in international comparative research, pointing to the cross-cultural comparability of the construct and its effects (e.g. OECD, 2011; Martin et al. 2008; Schmidt et al. 2001). Yet although there is an overall positive relation of learning time and achievement, there are large differences within and between countries and between different groups of students or schools (Ghuman and Lloyd, 2010; OECD, 2011).

Overall it is important to distinguish that learning time is *provided* by the school system, *realized or implemented* by the school and the teacher in the classroom, and *used* by the students. On this path from "gross" learning time as allocated in system level policies to student "time-on-(the right) task", many factors on four different levels (system, school, classroom, and student) diminish available learning time to a different degree across countries (Gillies and Quijada, 2008; Benavot, 2004). Differences in the amount of loss for various subgroups of students is an indicator of equity in educational opportunities, because research shows that relations with outcomes are stronger when learning time is more narrowly defined (e g., time-on-task instead of allocated learning time). Therefore, PISA 2015 intends to apply a broader view on learning time (Abadzi, 2009, cf. Table 5; see also Berliner, 1990, Millot and Lane, 2002).

On system level PISA 2015 assesses *allocated learning time (at)* the time which a school system intends to educate its students (e.g. “number of school weeks per year” multiplied by “number of school hours per week”). This sets the outer limit of time that theoretically could be used for teaching and learning. On school level, this has to be *provided (pt)*, but there is loss on national level (such as weather, holidays, or aggressive conflicts) and on school level (e.g. local festivities, teacher strikes, illness, or other teacher absenteeism; cf. Ghuman and Lloyd, 2010; Chaudhury et al. 2006). Another proportion of time is then dispensed in the classroom, resulting in *realised learning time (rt)*. Time loss at classroom level is most commonly due to non-teaching activities like classroom management, collecting homework, or waiting time (e.g. MacCay, 2009; The PROBE Team, 1999) and leaves realized learning time as the fraction of time during which a class is taught.[7] The proportion of realised learning time during which a student actively attends to learning content is *engaged learning time (et)*. This excludes periods in which a student does not attend due to illness, truancy, being late, or being present but absent-minded. Engaged learning time is the only time during which students actually learn. Ultimately, a certain fraction of time is dedicated to curriculum-related learning content and therefore called *time-on-task (tot)*. It accounts for the proportion of overall learning time which is most proximal to a student’s learning process and therefore to student outcome.

Moreover, it has been shown that next to the absolute amount of time available for learning, students’ time-use patterns relate to important success variables and can help explain relations between student background variables (such as ESCS) and performance variables (e.g. as mediator variables, cf., Porterfield and Winkler, 2007) Also, patterns of free-time activities in middle childhood predict adjustment in early adolescence (McHale, Crouter and Tucker, 2001) and participation in extracurricular activities has been demonstrated to protect at risk students from early school drop-out (Mahoney and Cairns, 1997) making a point to include a broader view on student’s time use but on learning time in school alone. In the PISA 2015 FT, students’ time use before and after school will be assessed with a set of newly designed questions that was developed in parallel to Kahneman’s et al. (2004) “Day-reconstruction method”.

[7] It is important to note that at least at this level “time loss” refers to diminished learning time that focuses on curricular content and therefore on domain-specific cognitive outcome. More overarching goals of education such as self-regulation, interest or social competencies might very well be stimulated during “lost” time periods.

Table 5: Assessment of Learning time and loss of learning time in the PISA 2015 FT

		Student Questionnaire	**School Questionnaire**	
use	**student**	- Reasons for loss - Amount of loss		Time-on-task (tot) = ET - non-subject related engaged time
use	**student**	+ Additional instruction and study (time use) - Truancy - Reasons for loss		Engaged time (ET) = RT - student absenteeism, truancy, absent minded time
provision	**classroom**	- Disciplinary climate and loss in science classes		Realized learning time (RT) = PT - loss due to classroom management, assessment time, waiting time, etc.
provision	**school**	+ Amount of school learning time + Number and type of science classes	- Loss on school level Policy and practices on additional instruction Other study support	Provided learning time (PT) = AT - loss due to weather, holidays, teacher absenteeism, etc.
provision	**system**		+ Instructional days - Loss on system or regional level	Allocated learning time (AT) = Total/ learning time: duration x dosage

Also, besides learning times for mandatory schooling, additional in-school and out-of-school learning activities are taken into account. PISA 2015 attempts to identifying additional learning time in a cross-culturally valid way, incorporating for example different formats, location, content, and purposes. Information from the School and Student Questionnaire, and from the optional Educational Career Questionnaire—can be combined to get the full picture. Similarly, information on informal, extra-curricular learning activities and time use before and after school every day as well as science-related experiences is gathered from students, from parents in the optional Parent Questionnaire, and from school leaders in the School Questionnaire (cf. also Module 5).

Curriculum: The school curriculum embraces all programmes and learning content that students are assigned to. Great differences can exist between system level intentions on curriculum, the implementation at teacher level or in textbooks, and the curriculum that is actually experienced by the students. For the major domain of PISA 2015, “science”, differences in curriculum are particularly large across tracks, grades, schools, and countries (Schmidt et al. 2001; Martin et al. 2008). Thus the variety of possible curriculum content aspects to be assessed (particularly in science) cannot be covered in the context questionnaires. Instead, they focus on internationally comparable aspects of curricula (Schmidt et al. 2001; Schmidt et al. 1999; Schmidt, Wang and McKnight, 2005): coherence (alignment of learning content within and across units, grades, subjects, and disciplines), focus (number of topics covered in a given time span), rigour (performance and curriculum coverage expectations).

International curriculum research has found evidence for effects on learning outcomes for all three aspects (Schmidt and Houang, 2007: relations of greater curriculum incoherence to poorer student performance; Gamoran et al. 1997: positive relations between achievement and alignment of content coverage; Schmidt et al. 2001: students in countries teaching fewer science topics have higher average achievement levels; Schmidt et al. 2001: content standards are positively related to student learning gain). However, most studies only regard curriculum provision (i.e. national guidelines, teacher diaries, textbook content coverage) while few take into account the student perspective, engagement or participation (e.g. in private additional learning activities).

Table 6: Assessment of the science curriculum in the PISA 2015 FT

	School and system level: intended curriculum	**Classroom and student level: implemented curriculum**
coherence	Implementation of standards for science teaching (ScQ) Curriculum in science/computer use (ScQ) *Components of the Science Curriculum (TcQ Science)*	Implementation of a common science curriculum (ScQ) Science classes taken this year and last year (StQ)
focus	Student achievement and non-curricular competencies are subject to evaluation (item in ScQ) *Components of the Science Curriculum (TcQ Science)* *Content of the science curriculum (TcQ Science)*	Science teaching activities (StQ) Freedom in choosing science classes (StQ)
rigour	*Existence of a common science curriculum (TcQ Science)* *Content of the science curriculum (TcQ Science)* *Availability of the science curriculum to teachers, parents, and students (TcQ Science)*	Implementation of the curriculum is subject to evaluation (item in ScQ) Science classes taken this year and last year (StQ) Science teaching activities (StQ)

Note: Constructs in italics are assessed in the optional Teacher Questionnaire.

The curriculum is provided or intended at system and at school level, delivered or implemented in schools at classroom or teacher level, and experienced and learned by individual students. Every stakeholder thereby perceives a certain degree of coherence and focus and holds certain expectations (Travers and Westbury, 1989). At system level, there is an intention of what students should learn and how they should learn it, the *intended curriculum*. On school level, the faculty, the principal, or the science coordinator transfer and interpret the intended curriculum. This might be done by choosing and emphasizing specific topics, textbooks, and educational goals. Together with the curriculum realized on teacher level, this is the *implemented curriculum*. In the classroom, the teacher implements the curriculum, aligns teaching methods and topics, and supports students in transferring knowledge and skills across topics and grades. Even in countries with a fixed mandatory, centralised curriculum, variation exists in the different aspects of curriculum at the different levels (Rowan, Camburn and Correnti, 2004; Schmidt et al. 1999).

PISA 2015 therefore includes curriculum aspects, coherence, focus, and rigour at intended and implemented level to study the science curriculum. The School Questionnaire (ScQ), the Student Questionnaire (StQ) and the optional Science Teacher Questionnaire (TcQ Science) each contribute specific information (see Table 6).

2.3.3. Teacher qualification and knowledge/beliefs (Module 1)

Many studies have demonstrated a clear influence of teacher-related factors on student learning and outcomes. The growing focus on teacher-related policies within the OECD was exemplified by the Second Teacher Summit held in March 2012 (cf. Schleicher, 2012). In addition to teachers' professional behaviour within the classroom (see section 2.3.1 above), the composition of the teaching force in terms of, e.g. age and educational level, their initial education and qualification, their individual beliefs and competencies, as well as professional practices on the school level—such as collaboration and professional development—have been core topics in educational policy. Basic information on these topics will be available from the PISA 2015 Field Trial School and Student Questionnaires, while the optional Teacher Questionnaire—in part based on instruments previously established in the OECD Teaching and Learning International Study (TALIS)—features additional constructs, both science-specific and domain-general, as shown in Table 7.This instrument is new to PISA, although national instruments have been added to the PISA design successfully in Ireland and Germany in previous cycles, with broad support from the teaching force. Also, other Large Scale studies such as TIMSS have implemented teacher questionnaires without any loss in acceptance or participation. Thus, the new optional instrument will bring voice to the teachers so they can fully participate in PISA.

Table 7: Teacher-related measures in the PISA 2015 FT

<table>
<tr><th></th><th>Science-related</th><th>General</th></tr>
<tr><td>Background</td><td colspan="2">Gender, age, employment status, job experience, subjects studied, teaching modal grade?, workplace selection
Recruitment policies (ScQ)</td></tr>
<tr><td rowspan="2">Initial education</td><td colspan="2">Goal of first qualification, type and duration of teacher education and training programme (if attended), level and mode of qualification
Number of teachers by education level (ScQ)
School policies: Teacher Education or training programme Required (ScQ)</td></tr>
<tr><td>Science-related content
Number of science teachers by level of qualification and by subject (ScQ)</td><td></td></tr>
<tr><td rowspan="2">Professional development</td><td colspan="2">Participation in, and duration of different type of activities
Obligation and support
Amount of participation, duration, school policies (ScQ)</td></tr>
<tr><td>Collaboration
Science-related content</td><td>Cooperation
General content</td></tr>
<tr><td>Beliefs</td><td>Self-efficacy (related to science content and teaching science)
Performance vs. needs orientation
Enthusiasm (with anchoring vignettes)
Engagement (StQ)</td><td>Self-efficacy (general)
Job satisfaction</td></tr>
</table>

Note: If not indicated otherwise, constructs are included in the optional PISA 2015 Teacher Questionnaires.

Across these topics, a distinction is adapted that Shulman (1985) suggested for research on teachers: Teacher's beliefs and activities can be related either to the subject matter taught, its conceptual foundations, basic ideas etc. (content), or to teaching and learning the subject matter, including issues of student understanding, teaching practices, assessment procedures etc. (pedagogical content), or to general concepts such as classroom management (pedagogy).

Shulman's model has been most influential in research on teachers (e.g. Hill, Rowan and Ball, 2005; Baumert et al. 2010; Bloemeke et al. 2012). In line with this research, PISA 2015 identifies content, pedagogical content, and/or pedagogy as foci of most teacher-related constructs, including professional beliefs. It should be noted that there is no attempt to measure teacher knowledge.

Teacher background and initial education: Understanding the multiple pathways leading to the teaching profession—including career changes such as an engineer who now teaches science classes—is most important for educational policy, because there is a growing need to recruit teachers from non-traditional backgrounds. For these but also for novice teachers with traditional training, the stage of induction seems to be very important (Portner, 2005). Teacher retention is another concern in many countries (Ingersoll and Perda, 2010). In addition to formal qualifications (tertiary/secondary education certificates and academic degrees), a major in the subject being taught, type and duration of teacher education and training programmes attended, and professional experience (i.e. years having taught science at school), PISA 2015 asks teachers about the representation of the three foci in their initial education. Similarly, in asking about professional development, it is important to gather information on its focus in addition to frequency and type of professional learning.

Professional Development and Collaboration: Professional development refers to any activity that equips teachers with tools and resources necessary to provide quality instruction. It includes school-based programmes as well as networking, coaching, seminars, or other types of training activities that foster in-service learning and thus promote professionalization of teaching. Even though professional development is generally regarded as crucial for improving teaching and student achievement, Sykes (1996) referred to the ineffectiveness of common trainings as “the most serious unsolved problem for policy and practice” (p. 465). However, more recent studies report positive effects on teaching practices and classroom climate (Cuevas, Lee, Hart and Deaktor, 2005; Desimone, Porter, Garet, Yoon and Birman, 2002; Guskey, 2002; Jeanpierre, Oberhause, and Freeman, 2005; Supovitz and Turner, 2000; Timperley et al. 2007), as well as student achievement (e.g. McDowall, Cameron, Dingle, Gilmore and MacGibbon, 2007; Shayer and Adhami, 2007). This apparent inconsistency may be partly resolved by accounting for different features of the programmes examined. Summarising previous studies, Buczynski and Hansen (2010) describe ineffective programmes as being “too conventionally taught, too top–down, and too isolated from school and classroom realities to have much impact on practice” (p. 600). As early as in the 1980s, scholars indicated the benefits of supportive networks for teachers (e.g. Darling-Hammond, 1984; Rosenholtz, 1989; Bryk and Driscoll, 1988). In the 1990s the idea of the “professional learning communities” emerged. This notion refers to groups of teachers who co-operatively reflect and improve their professional practices (Hord, 1997). Research on professional learning communities is still limited, but there is some indication of positive effects on educational processes and outcomes (e.g. Lomos, Hofman, and Bosker, 2011). In China, for example, teachers are often organised in groups that work together studying national guidelines and defining teaching goals, that co-operate for preparing and improving teaching, and that organise observation visits to provide colleagues with feedback and involve teachers in out-of-school activities (Paine and Ma, 1993). Similarly, in Japan “lesson studies” are common practice among teachers (Stigler and Hiebert, 1999). The OECD Teaching and Learning International Survey (TALIS) further suggests that the pattern of activities also varies between countries (Vieluf, Kaplan, Klieme and Bayer, 2011).

The PISA School Questionnaire in 2000 and 2012 included a question about the proportion of teachers who had recently (within the previous three months) participated in any kind of professional development. In 2012, the same question was asked with a focus on mathematics teachers. However, this information did not show any substantial relation to student outcome (OECD, 2005). Therefore, PISA 2015 intends to enhance measurement of professional development by adapting questions from TALIS and other sources.

For one aspect of learning communities, namely teacher cooperation, a measurement approach has been developed that allows for a classification of schools into levels of cooperation, in line with a distinction between different levels of school development (Steinert, Klieme, Maag Merki, Döbrich, Halbheer and Kunz, 2006). In these models the lowest level of co-operation is restricted to a sporadic exchange of information and materials. Higher levels include increasing degrees of interaction, critical reflection, coordination, and trust. They further require a definition of common goals and a high frequency and systematisation of concerted action. The higher levels of co-operation also encompass most additional criteria of professional learning communities.

Professional beliefs: PISA takes up a measure of job satisfaction from TALIS. TALIS also measures another construct which is well established in educational research: self-efficacy beliefs. According to Tschannen-Moran and Woolfolk-Hoy (2001), three dimensions of teacher self-efficacy can be described using a shortened version of the question: Efficacy in classroom management, efficacy in instruction, and maintaining positive relations with students. These dimensions will be addressed in the domain-general TQ, while science teachers will be asked to report self-efficacy beliefs regarding science content and its teaching.

Teachers' morale and commitment was assessed in PISA cycles 2000, 2003 and 2012 in the School Questionnaire that was filled out by the principal (or some other member of the school management team), aiming to assess attitudes among teaching staff. These measures are taken up for PISA 2015 in Module 13, "school climate". The newly invented optional Teacher Questionnaire adds valid and reliable measures for two kinds of beliefs that are well-established in educational research. The data gained with such measures correlates positively with student learning, learners' engagement and willingness to learn (Patrick et al. 2003; Turner et al. 1998): performance vs. needs orientation (describing the strength of academic expectations), and teacher enthusiasm. Kunter et al. (2011) distinguish enthusiasm for teaching from enthusiasm for the subject. Pursuant to the framework outlined above, the scales serve as good models for assessing enthusiasm with respect to different foci. Finally, PISA 2015 aims at establishing a third perspective on teachers, defined by students' perceptions of teacher engagement: Students are asked to think about their recent experience in a specific science course, and judge their teacher's interest both in teaching their class and in the subject matter being taught.

The main level of analysis for data gathered in the optional Teacher Questionnaire is the school level. No weighting for individual teacher responses will be available. All data from the Teacher Questionnaire will therefore be basically treated as school variables.

2.4. Assessing school policies and governance[8]

This chapter summarizes the conceptual foundations for high priority module 19 (Assessment, Evaluation, and Accountability) as well as those of lower priority modules 3 (School-level learning environment for science) and 13-18.

2.4.1.Assessment, evaluation, and accountability (Module 19)

Assessing students and evaluating schools[9] is a common practice in most countries. Since the 1980s, policy instruments such as performance standards, standard-based assessment, annual reports on student progress, and school inspectorates have been promoted and implemented across continents. Reporting and sharing data from assessments and evaluations with different stakeholders provides multiple opportunities for monitoring, feedback, and improvement. In recent years, there has been a growing interest in the use

[8] This chapter is based on working papers submitted by Sonja Bayer, Eckhard Klieme, and Nina Jude for Module 19, Leonidas Kyriakides for Module 3, Silke Hertel, Nadine Zeidler, and Nina Jude for Module 14 (Parental Involvement) and Bieke de Fraine for Module 15 (School Management).

[9] The terms evaluation and assessment are defined quite differently in the literature. Sometimes they are even treated as synonyms. In the following we share the definition used in current OECD literature (see e.g. Rosenkvist, 2010): The term evaluation or school evaluation is used for processes on school and system level. Evaluators collect evidence to judge systems, educational programmes, policies and practices. This may include an evaluation of individual performance among professionals, such as teacher evaluation. Assessment or student assessment, on the other hand, directly refers to student performance or student learning processes (see also Harlen, 2007a). Notably, there is a strong link between assessment and evaluation. For instance, results from student assessments may be used for school evaluation purposes.

of assessment and evaluation results through feedback to students, parents, teachers, and schools as one of the most powerful tools for quality management and improvement (OECD 2010, p. 76). Accountability systems based on these instruments are increasingly common in OECD countries (Scheerens, 2002, p. 36). Accountability is oftentimes, but not necessarily linked to market-oriented reforms. Rewards and penalties for good and less good assessment and evaluation results are said to change behaviours in ways that improve student achievement (Woessmann et al. 2009). However, there are huge differences in assessment and evaluation practices and purposes[10].

Previous PISA cycles already covered aspects of assessment, evaluation and accountability in the School Questionnaire, with a strong focus on the use of standardised tests. In PISA 2015 this module asks both about standardised and less standardised practices. Internal and external evaluations address different purposes and consequences and will be dealt with separately. Also, teacher evaluation and teacher incentives are addressed as means of quality management. Finally, formative assessment and feedback are increasingly popular in research and teaching practice. For all four types of assessment and evaluation, their respective purpose and criteria, practices, use and consequences are addressed (Pellegrino, Chudowsky and Glaser, 2001; Scriven, 2003; Wilson, 2004; see Table 8). All constructs, with the exception of classroom assessment and grading, are covered by the standard School Questionnaire (ScQ) or Student Questionnaire (StQ). However, The questions on teacher evaluation and internal evaluation are also included in the optional General Teacher Questionnaire (TQG) or the Science Teacher Questionnaire (TQS) in order to cross-check this information from a different perspective.

In the following, relevant research on school evaluation and student assessment is summarised to provide the rationale for questionnaire development in PISA 2015.

Evaluation: The evaluation of schools is used as a means of assuring transparency, deciding and making judgements about systems, programmes, educational resources and processes, and also to guide school development (Faubert, 2009). Evaluation criteria need to be defined and applied from the viewpoints of different stakeholders (Sanders and Davidson, 2003).

Evaluation can be either external or internal (Berkemeyer and Müller, 2010). It is called external evaluation if the process is controlled and headed by an external body and the school does not define the areas which are judged. An evaluation is called internal if it is part of a process controlled by the school and in which the school defines which areas are judged; the evaluation may be conducted by members of the school (self-evaluation) or by persons/institutions commissioned by the school. Different evaluation practices generally coexist and benefit from each other (Ryan, Chandler and Samuels, 2007). External evaluation can expand the scope of internal evaluation, also validate results and implement standards or goals. Internal evaluation can improve the interpretation and increase the utilisation of external evaluation results (Nevo, 2002). However, improvement of schools seems to be more likely when an internal evaluation was applied, compared to external evaluation. Therefore processes and outcomes of evaluation might differ between internal and external evaluation. Moreover, country and school specific context factors may influence the implementation of evaluations as well as conclusions and effects for schools. In many countries, individual evaluation of teachers and principals, separate from school-wide evaluation, is also common (Faubert, 2009; Santiago and Benavides, 2009); they are treated here as a separate type of evaluation.

[10] See OECD's country reviews of Assessment and Evaluation http://www.oecd.org/education/preschoolandschool/oecdreviewonevaluationandassessmentframeworksforimprovingschooloutcomescountryreviews.htm.

Table 8: Measures in the PISA 2015 FT related to assessment, evaluation and accountability

	External evaluation	**Teacher evaluation**	**Internal Evaluation**	**Formative assessment**
Purpose and criteria	General assessment practice (ScQ) Purpose of assessment results (ScQ)			
	Evaluation policies (ScQ)		Foci of internal evaluation (ScQ, TQG)	Teacher's grading (TQG)
Practices	Processes of external evaluation (ScQ)	Teacher evaluation methods (ScQ, TQG)	Processes of internal evaluation (ScQ, TQG)	Classroom assessment instruments (TQG/TALIS)
Use and consequences	Use of achievement data for accountability (ScQ)	Teacher incentives (ScQ, TQG)	Consequences of internal evaluation (ScQ, TQG)	Feedback: student perception (StQ). Use of feedback to guide learning (StQ) Adaptation of instruction (StQ,TQS)

Results of evaluations may be used in a formative way (e.g. to guide the analysis and improvement of processes) or in a more summative way (e.g. for accountability). Formative evaluation aims at closing the gap between the as-is state and the target state. Here, teaching and school-based processes are to be guided to a predetermined goal. Summative evaluation focuses on student outcomes and encourages schools to meet specific standards. Formative evaluation has turned out to be more effective in school improvement than summative evaluation (Creemers and Kyriakides, 2008). Effects or consequences of evaluation may differ, depending on e.g. the focus of evaluation, the procedure chosen for the evaluation, or a school's goals and priorities.

Assessment: Communication and clarification of achievement goals within schools is essential in students' learning (Brookhart, 2007; Stiggins, 2007). National standards that have emerged in recent years define what students should know (Koeppen et al. 2008; Shepard, 2006). These educational standards directly shape school policies and classroom instruction by urging schools and teachers to fine-grain the standards and communicate specific aims leading to a shared understanding. To check whether these goals are met, schools follow a given assessment practice or define their own. This can be implemented in the classroom learning process by more or less standardised tests and oral examinations developed by the teacher. Additionally, mandatory and non-mandatory standardised and externally developed tests verify and compare student outcomes on classroom, school, district, state or international level (Shepard, 2006). Irrespective of the purpose and the stakeholder administering the assessment, a test must fulfil a number of quality criteria (Scheerens, Glas and Thomas, 2003). In general, standardised tests are more reliable measures but may be less aligned with the school curriculum, and vice versa for teacher-made assessments.

The distinction between formative and summative, internal and external approaches also holds for student assessment. In its summarising function, assessment takes place in order to grade, certify, or record progress. Summative assessment, whether external or internal, therefore indicates and monitors standards, but may also raise standards by causing students as well as teachers and schools to put more effort into their work (Harlen and Deakin Crick, 2002). On the other hand, summative assessment might lead to lower self-esteem and diminished effort of students at risk and therefore can increase the gap between lower and higher achieving students (Black and Wiliam, 2004). Another negative aspect of assessment may arise when teaching solely focusses on answering questions, rather than on developing skills and knowledge (Harlen and Deakin Crick, 2002). Grading is arguably the most prevalent type of classroom assessment, and an essential aspect of effective teaching (McMillan, 2001; Guskey, 2007). Grades have been shown to be unreliable and of limited validity, however, but there is very little comparative research on grading practices in different countries.

Formative assessment can be a significant source of improvement in student learning processes (e.g. Shepard, 2006; Black and Wiliam, 2004; McMillan, 2007, OECD, 2005a). Especially for low achievers, formative assessment can lead to sizable gains in student achievement (Abrams, 2007). Formative assessment and reciprocal feedback might not just be useful for students, but also for teachers, helping them to adapt their instruction to their students' needs. However, there is large variation in the implementation and in the impact of formative assessment practices (e.g. Kingston and Nash, 2011; Shute, 2008; Hattie and Timperley, 2007; Black and Wiliam, 1998). Therefore, it is worthwhile to study cross-national variation in formative assessment practices through PISA 2015.

The School Questionnaire for PISA 2015 will take up several questions on general assessment practices and results, external evaluation, and teacher evaluation, which have been used in previous cycles, to report trends. Information on teacher evaluation is asked in parallel in the Teacher Questionnaire in order to compare and cross-validate information from different stakeholders. However, in line with the research cited above, internal school evaluation and formative, classroom-based assessment have been given more weight in PISA 2015 than in previous cycles.

2.4.2. Other school policies and approaches to educational governance

During the last two decades, research on educational effectiveness has largely been concerned with the impact of school-level factors on students' learning. Studies show that school qualities have effects on student progress, with variation in schools appearing to affect students' behaviour. It has been asserted that the environment at the school level can influence the behaviour of teachers and students and thus—mostly indirectly—their consequent success in teaching and in learning. Both "soft" factors such as school climate and parental involvement, and "hard" factors such as school management activities and allocation policies vary and are related to student outcomes within and across countries.

School climate (Module 13): School climate encompasses shared norms and values, the quality of relationships and the general atmosphere of a school. An academic focus—a general consensus about the mission of the school and the value of education, shared by school leaders, staff, and parents—impacts the norms in student peer groups, and facilitates learning. In addition, an orderly learning atmosphere maximises the use of learning time. By contrast, disrespectfulness and an unruly environment are counterproductive for teachers and students alike and distract from the school's actual mission. As in previous PISA assessments, school climate will be assessed in the Student Questionnaire ("student-teacher relationship", "achievement pressure") and the School Questionnaire ("teacher morale", "behaviour affecting school climate").

However, a number of items have been added to the teacher-student-relationship scale in order to cover different aspects of supportive learning environments that have been identified by Caldwell and Bradley (1984) in their "Home Inventory": emotional warmth, guidance and boundaries, stimulation/scaffolding and stability. The complete question has been implemented several times in the PISA 2015 FT: students report on their relationship with teachers and parents, while teachers and parents (in the optional questionnaires) will be asked to answer parallel items. All in all, a complex picture of social relationships between students, teachers, and parents could emerge in countries which administer both the parent and the teacher questionnaires. Finally, two scales have been added to the Student Questionnaire that cover more problematic, oftentimes hidden aspects of school climate, which nevertheless are highly important from a pedagogical as well as a policy point of view: bullying by peers and unfair treatment by teachers. If the Field Trial results provide evidence of trustworthy results, the PISA description of school contexts may become much richer in the future.

Parental involvement (Module 14): Over the past years, the involvement of parents in educational processes has gained importance in the educational debate, and to some extent it has also become relevant for educational policy. Parents are not only an important audience, but powerful stakeholders in education. Thus information on parents' opinions and engagement is highly valuable also for large scale assessments like PISA. Parental involvement in education has been part of PISA since 2006 when the parent questionnaire was administered for the first time, directly addressing the parents of the PISA students. For PISA 2015, specific aspects of parental involvement were added to all questionnaires, focussing on a) parent-school communication and collaboration, and b) specific aspects of parental support for learning. Depending on the addressee of the questionnaires, i.e. teachers, students, school administrators, or parents, specific aspects of parental engagement are highlighted. Aspects of parental involvement can thus be found within scales related to other modules. Nevertheless, the majority of items and topics regarding parental involvement are included in the parent questionnaire that will be administered as an international option.

Leadership and school management (Module 15): School principals play a key role in school management. They can shape teachers' professional development, define the school's educational goals, ensure that instructional practice is directed towards achieving these goals, suggest modifications to improve teaching practices, and help solve problems that may arise within the classroom or among teachers. The PISA measure of educational leadership saw a major revision in the 2012 study. This work has been reviewed, and the leadership scale could be considerably shortened. In addition, the new optional Teacher Questionnaire will be used to gather information on instructional and transformational leadership from teachers as well, because research has shown that the teachers' perspective on leadership can differ from the positions held by school administrators.

Resources (Module 16): Information on school type (public vs. private) and class size has always been included in the School Questionnaire. In addition to these trend questions, PISA 2015 allows to discriminate between types of private schools (religious/denominational, not-for-profit, for-profit) and provides more advanced information on ICT use. All PISA cycles so far have included a question on the degree of problems a school experiences due to missing resources. The different approaches over time were systematized and implemented in one coherent question in the School Questionnaire.

Locus of decision making (Module 17): Educational systems have been classified by the amount of control that is left to the school (i.e. school board, staff, and school leaders) when decisions on admission, curriculum, allocation of resources, and personnel have to be made. These indicators are based on questions in the School Questionnaire which are left unchanged to allow for trend reporting.

Allocation, selection, choice, and grade repetition (Module 18): The way students are channelled into educational pathways, schools, tracks, or courses, is a core issue of educational governance ("stratification"). On the school level, selection and allocation procedures are important aspects of school organisation. Highly selective schools provide a learning environment that may differ from the environment offered by more comprehensive schools. For all those reasons, appropriate trend questions answered by school administrators and parents have been kept.

As a new topic, PISA 2015 adds information on grade repetition, which is a very important aspect of vertical differentiation, as shown in the PISA 2009 report. Many longitudinal studies have demonstrated grade retention to have a negative impact on individual careers and outcomes (e.g. Ou and Reynolds, 2010; Griffith et al. 2010), student behaviour and well-being (e.g. Crothers et al. 2010). Grade repetition is less common in secondary schools compared to primary schools, but late retention seems to have larger (negative) effects (Ou and Reynolds, 2010). Greene and Winter (2009) showed that once a test-based retention policy has been installed, those who are exempted from the policy do worse. Babcock and Bedard (2011) show that larger numbers of grade retention may even have a positive (!) effect on the cohort, i.e. all students, including those who are promoted. Kloosterman and de Graaf (2010) argued that in highly tracked systems, such as in some European countries, grade repetition might serve as a preferred alternative to changing into a lower track; indeed, they find evidence that this strategy is preferred for students with higher SES. Thus, changing grade repetition policies might be a viable option regarding low-cost interventions (Binder, 2009). Therefore, it is worthwhile to take a closer, comparative look at grade retention policies and their (differential) effects on students who do, and do not repeat a grade. PISA 2015 will explore two related questions in the School Questionnaire and includes a question on grade repetitions in the Standard Student Questionnaire.

School learning environment for science (Module 3): Conceptually, this module overlaps to a considerable degree with other modules dealing with school-level factors, such as Module 12, Learning Time and Curriculum, Module 15, School Leadership and Management, or Module 19, Assessment, Evaluation, and Accountability (cf. above). In addition to those, the QEG requested to include some questions in the School Questionnaire that would directly focus on the status of science education in the school and available resources. Accordingly, a question has been developed to investigate the overall value of science within the school (Is science education a shared priority for stakeholders?), along with questions on resources particularly available for science education: size of teaching staff per science area; resources such as laboratories and equipment for student experiments; cooperation with science museums, companies, and other external partners.

2.5. Assessing student background (Modules 7 – 9)[11]

This chapter covers three modules that were given lower priority by the PISA Governing Board: module 7 (Student SES, family and home background), module 8 (ethnicity and migration), and module 9 (educational pathways in early childhood). Nevertheless, these topics, and module 7 in particular, require careful revisiting, because they contain the basic information needed for calculation of the Index of Economic, Social and Cultural Status (ESCS).

Student SES, family and home background (Module 7): In order to compare equity related to social and ethnic factors across PISA cycles, the PISA consortium (Core 6) staff intend to keep measures of socio-economic status and other background variables basically unchanged. However, some minor changes have become necessary. Due to extensive development in the ICT sector for example, questions on technology equipment in the student's home were slightly outdated. Thus, the measures of home possessions will be updated to ensure better coverage of within and across country variation of home possessions. These changes are expected not to impact the important trend measures in this module. Additionally some experimental measures will be included to capture a more detailed picture of the home setting, its economic, educational, and social capital (Bourdieu, 1986; e.g. source of funding for educational expenses, communication with parents, parents' interaction with student's friends, cultural activities, etc.). Because the indicators derived from questions related to Module 7 are most important for later reporting purposes but also for data management and preparation, further measures have been implemented to ensure that all necessary information will be available in the new PISA design and computer-based assessment mode. Most important, alternative ways of asking students about parental

[11] This chapter is based on working papers submitted by Wolfram Schulz for Module 7, Svenja Vieluf for Module 8, Susanne Kuger and Hans-Günter Roßbach for Module 9.

occupations in the Field Trial have been implemented, because the quality of these data may be endangered when the answering mode for the Student Questionnaire changes from writing with paper and pencil in previous cycles to typing on a computer. To this end two differently complex systems of assessment were adapted from other international surveys.

Ethnicity and Migration (Module 8): Linguistic and cultural diversity are basic facts of life in most regions of the world. Many nations are home to several subpopulations with different languages and cultures. International migration perpetuates this diversity. In OECD countries, first and second generation immigrant students currently comprise 10 to 20% of the student population (OECD, 2010a; 2010b). At the same time, students from ethnic minority groups and immigrant students often face particular challenges. In a number of education systems, immigrant students perform at significantly lower levels than their native peers in key school subjects (Stanat and Christensen, 2006), and both groups are often faced with overt or covert discrimination with potentially detrimental consequences for their psychological development and well-being. Thus, providing students from different linguistic and cultural backgrounds with equal opportunities is often considered one of the central challenges for education systems in the 21st century (e.g. OECD, 2010a).

PISA 2015 puts a special focus on diversity-related aspects of the school climate. A new question developed for the PISA 2015 Field Trial asks students about their membership in a group that they believe to be discriminated against in their country. If they identify themselves as belonging to such a minority group, they are asked whether they feel treated in a respectful and fair manner by their teachers and equal to their classmates– a factor which has been shown to be related with educational outcomes (e.g. Fisher, Wallace, and Fenton, 2000; Wong, Eccles, and Sameroff, 2003). Another new question, implemented in the optional Parent Questionnaire, assesses perceived barriers to parental involvement. Additionally, teachers and principals are asked about diversity-related assumptions among teachers in their school. The wording of the question is based on research on multiculturalism support (Van de Vijver, Breugelmanns, and Schalk-Soekar, 2008). In PISA 2015, forced choice items are used to contrast the two poles of this dimension. Additionally, PISA 2015 examines palpable aspects of multicultural education practices, or the extent to which multicultural education practices are implemented in different schools. Altogether, findings from this module may help to better understand educational inequalities and can stimulate teacher training programs, school development activities focusing on diversity management strategies, or policy interventions supporting parents from diverse populations to get more involved in their children’s schooling.

Educational pathways in early childhood (Module 9): When children enter primary school, they already differ in their language, pre-reading, and early numeracy skills and these differences are often maintained later in life. Promoting school readiness and better adjustment to school is hypothesised to be an efficient means of raising the achievement levels of all children, but especially of those children who experience a lack of parental support or who grow up in disadvantaged circumstances. It has been argued that investing in early education programmes will have large long-term monetary and non-monetary benefits (Heckman, 2006). The importance of pre-school quality has been acknowledged and analysed by OECD reporting as well.

According to UNESCO (2006), **E**arly **C**hildhood **C**are and **E**ducation is defined as “programmes that, in addition to providing children with care, offer a structured and purposeful set of learning activities either in formal institutions (pre-primary) or as part of a non-formal child development programme”. The focus of the international comparable statistics International Standard Classification of Education Level 0 (ISCED 0) is much narrower. Currently at least four strands of research support the relevance of applying a broader definition of ECCE than focussing on ISCED 0 alone: brain research, studies on domain-specific development and support, evaluation studies of model programmes, longitudinal large-scale studies all rely on the broader definition of ECCE. Thus, conclusions about the importance of early child care should be drawn with ECCE and not with ISCED 0 in mind.

However, when evaluating the body of research it becomes obvious that in fact, a number of characteristics of *the kind of ECCE provided* seem to determine whether benefits can be observed or not and whether these benefits disappear or persist. Students' early childhood opportunities to learn are best assessed in terms of curriculum, quantity and quality of early childhood learning experiences. For example, one of the best sources available, the British EPPE Study, did find short-term effects showing that pre-school attendance was beneficial for cognitive and socio-emotional development in particular for children from disadvantaged backgrounds. However, in the long term only those children showed persisting beneficial pre-school effects who attended a high-quality pre-school centre (e.g. Sammons et al. 2008; Sylva, Melhuish et al. 2011; cf also Valenti and Tracey, 2009). Also, a certain degree of intensity in terms of hours per week/months seems to be a precondition for beneficial effects of ECCE attendance (Logan et al. 2011; Sylva, Stein et al. 2011).

Thus, asking about early education experience in PISA only makes sense if specific aspects of dosage, quality, and curriculum can be retrieved retrospectively, which is more than unlikely (Fivush and Hamond, 1990; Markowitsch and Welzer, 2009). As a consequence, PISA 2015, while keeping a short question on ISCED 0 attendance in the Student Questionnaire, implemented a series of questions in the parent questionnaire, expecting parents to be the more reliable source of information. Those countries applying the optional parent questionnaire will gain information on basic characteristics of the early childhood education and care arrangements of PISA participants and reasons for attending or not attending early childhood education and care.

2.6. Next steps in questionnaire development

Questionnaire development for the PISA 2015 Field Trial aimed at a broad coverage of policy issues, organized by 19 modules which were solidly anchored in state-of-the art research on educational policy and educational effectiveness. Particular emphasis was given to modules that were judged as high priority by the participating countries. In search for innovation and optimization, multiple measures were implemented for some of the constructs. The Field Trial enables experimenting with different question formats—e.g. using interactive features of the computer-based administration system—and with new content—e.g. measures of students' physical and emotional well-being. The Field Trial will assess psychometric properties, but will also check whether new measures are accepted by stakeholders in different cultures. Based on the Field Trial findings, a subset of measures will be selected for the Main Study.

The core content defined in the first part of this Framework (section 1.3, Table 2) will take approximately two thirds of the Student Questionnaire (StQ) and less than half of the School Questionnaire (ScQ) in the PISA 2015 Main Study. This leaves some, albeit limited, space for additional content. All measures will be reviewed according to the following principles:

- Questions and items with proven validity and measurement quality will be given highest priority.
- Questions that can be shown to work well across cultures, and/or may help to describe country/culture differences, will receive priority.
- Questions that serve the needs of multiple modules are preferred for reasons of efficiency.
- If there are two or more similar constructs, all but one will be dropped.
- If a construct is assessed from two different perspectives (such as school policies related to evaluation, addressed in parallel in the School Questionnaire and in the Teacher Questionnaire), the perspective which has better psychometric quality will be retained.
- Short questions will be preferred to longer questions. Whenever possible, the number of items per question will be reduced based on Field Trial results.

- Quality otherwise being equal, constructs will be retained that have a stronger relation to student assessment results.

References

Abadzi, H. (2009), “Instructional Time Loss in Developing Countries: Concepts, Measurement, and Implications”. The World Bank Research Observer, 24(2), pp. 267–290.

Abedi, J., M. Courtney, S. Leon, J.C. Kao and T. Azzam (2006), *English Language Learners and Math Achievement: A Study of Opportunity to Learn and Language Accommodation (CSE Tech. Rep. No. 702)*, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California, Los Angeles.

Abrams, L.M. (2007), “Implications of Migh-Stakes Testing for the Use of Formative Classroom Assessment”, in J.H. McMillan (ed.), *Formative Classroom Assessment*, *Theory into practice*, Teachers College Columbia University, New York, pp. 79–98.

Ainley, J., E. Frances, C. Freeman, C. and K.O'Malley (2010), *ICT in the Teaching of Science and Mathematics in Year 8 in Australia: report from the IEA Second International Technology in Education Study (SITES) survey*, Melbourne: ACER. Retrieved from http://research.acer.edu.au/acer_monographs/6

Almlund M., A.L.Duckworth, J.J.Heckman and T. Kautz (2011), "Personality Psychology and Economics," *IZA Discussion Papers 5500*, Institute for the Study of Labour (IZA), Bonn.

Babcock, P. and K. Bedard (2011), “The Wages of Failure: New Evidence on School Retention and Long-Run Outcomes. Education Finance and Policy”, *Education Finance and Policy*, Vol. 6, No. 3, pp. 293–322.

Baumert J., M. Kunter, W. Blum, M. Brunner, T. Voss, A. Jordan et al. (2010), “Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress”, *American Educational Research Journal*, Vol. 47, No. 1, pp. 133–180.

Blömeke, S., U. Suhl, G. Kaiser and M. Döhrmann (2012), “Family Background, Entry Selectivity and Opportunities to Learn: What Matters in Primary Teacher Education? An international comparison of fifteen countries”, *Teaching and Teacher Education*, Vol. 28(1), pp. 44–55.

Benavot, A. (2004*), A Global Study of Intended Instructional Time and Official School Curricula, 1980-2000*, Background Paper commissioned by the International Bureau of Education for the UNESCO-EFA Global Monitoring Report 2005, The Quality Imperative, Geneva.

Berkenmeyer, N. and S. Müller (2010), „Schulinterne Evaluation - nur ein Instrument zur Selbststeuerung von Schulen?“ In: H. Altrichter and K. Maag Merki (Eds.): *Handbuch neue Steuerung im Schulsystem.* Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 195-218.

Berliner, D.C. (1990), “What's all the Fuss about Instructional Time?” in M. BenPeretz and R. Bromme (eds.), *Life in classrooms*, reissued with a new introduction, Teachers College Press, New York, pp. 3-35.

Binder, M. (2009), “Why Are Some Low-Income Countries Better at Providing Secondary Education?”, *Comparative Education Review*, Vol. 53, No. 4, pp. 513–534.

Black, P. and D. Wiliam (1998), “Assessment and Classroom Learning. Assessment in Education: Principles, Policy & Practice”, *Assessment in Education: Principles, Policy & Practice*, Vol. 5, No. 1, pp. 7–74.

Black, P. and D. Wiliam (2004), “Classroom Assessment is not (necessarily) Formative Assessment (and vice-versa),”, in M. Wilson (ed.), *Towards coherence between classroom assessment and accountability, Yearbook of the National Society for the Study of Education*, University of Chicago Press, Chicago, Ill., pp. 183–188.

Blanchard, M.R., S.A. Southerland, J.W. Osborne, V.D. Sampson, L.A. Annetta and E.M. Granger (2010), “Is Inquiry Possible in Light of Accountability? A Quantitative Comparison of the Relative Effectiveness of Guided Inquiry and Verification Laboratory Instruction”, *Science Education*, 94(4), pp. 577–616.

Blau D. and J. Curie (2006), "Pre-School, Day Care, and After-School Care: Who's Minding the Kids?" in *Handbook of the Economics of Education*, Elsevier, North Holland, pp. 1164-1278.

Bonita, R., M. De Counter, T. Dwyer, K. Jamrozi, and R. Winkelmann (2001), *Surveillance of Risk Factors for Noncommunicable Disease: The WHO STEPwise Approach*, World Health Organizatio, Geneva.

Bourdieu, Pierre (1986). *The Forms of Capital*. In J. Richardson (Ed.) Handbook of Theory and Research for the Sociology of Education (New York, Greenwood), 241-258.

Caldwell, Bettye M.; Bradley, Robert H. (1984): *Home Observation for Measurement of the Environment: Administration Manual, Revised Edition.* Unpublished manuscript. University of Arkansas. Little Rock, AK.

Brookhart, S.M. (2007), “Expanding Views about Formative Classroom Assessments: A Review of the Literature”, in J.H. McMillan (ed.), Formative Classroom Assessment, Teachers College Press, New York, pp. 43–62. Bryk, A.S., P. Sebring, E. Allensworth, S. Luppescu and J. Easton (2010), *Organizing School for Improvement*, Lessons from Chicago, The University of Chicago.

Bryk, A.S., and M. E. Driscoll (1988), *The High School as Community: Contextual Influences and Consequences for Students and Teachers*, Madison, Wisconsin: National Center on Effective Secondary Schools, University of Wisconsin.

Buczynski, S. and B. Hansen (2010). Impact of professional development on teacherpractice: Uncovering connections. Teaching and Teacher Education, 26(3), 599-607.

Chaudhury, N., J. Hammer, M. Kremer, K. Muralidharan and F. Rogers (2006), “Missing in Action: Teacher and Health Worker Absence in Developing Countries”, *Journal Of Economic Perspectives*, 20(1), 91-A4.

Conley, A.M., Pintrich, P.R., Vekiri, I., and D. Harrison, (2004), “Changes in Epistemological Beliefs in Elementary Science Students: Epistemological Development and its Impact on Cognition in Academic Domains”, *Contemporary Educational Psychology, 29*(2), pp. 186–204.

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.

Creemers, B.P.M. and G.J. Reezigt (1997), “School effectiveness and school improvement: Sustaining links”, *School Effectiveness and School Improvement, 8* (4), pp. 396-429.

Creemers, B.P.M., L. Kyriakides and P. Sammons (2010), *Methodological Advances in Educational Effectiveness Research*, Taylor & Francis, London and New York.

Cunha, F., J.J. Heckman, L.J. Lochner and D.V. Masterov (2006), "Interpreting the Evidence on Life Cycle Skill Formation,", in E. Hanushek and F. Welch (eds.), *Handbook of the Economics of Education*, Elsevier, North Holland, pp. 697-812.

Cuevas, P., O. Lee, J. Hart, and R. Deaktor (2005), “Improving science inquiry with elementary students of diverse backgrounds”, *Journal of Research in Science Teaching*, 42(3), pp. 337–357.

Darling-Hammond, L. (1984), *Beyond the commission reports: The coming crisis in teaching.* Santa Monica, CA: Rand Corporation.

Deci, E.L. and R. M. Ryan (1985*), Intrinsic motivation and self-determination in human behavior.* New York: Plenum Press.

Deci, E.L. and R. M. Ryan (1993), Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. *Zeitschrift für Pädagogik, 39*, pp. 223-238.

Desimone, L.M., A.C. Porter, M.S. Garet, K.S. Yoon and B.F. Birman (2002), "Effects of Professional Development on Teachers' Instruction: Results from a Three-Year Longitudinal Study", Educational Evaluation and Policy Analysis, Vol. 24 No. 2, pp. 81-112. Diener, E. and T. William (2006), *National Accounts of Well-Being,* Draft prepared for Encyclopedia of Quality of Life, p. 28.

Eccles, J. S. (1994), "Understanding Women's Educational and Occupational Choices: Applying the Eccles et al. Model of Achievement-Related Choices, *Psychology of Women Quarterly,* 18(4), pp. 585–609.

Eccles, J.S. and A. Wigfield (1995), In the Mind of the Actor: The Structure of Adolescents' Achievement Task Values and Expectancy-Related Beliefs,Personality and Social Psychology Bulletin, 21(3), pp. 215–225.

European Commission (ed.) (2004), Europe Needs More Scientists!, European Commission, Directorate-General for Research, High Level Group on Human Resources for Science and Technology in Europe, Brussels, *http://ec.europa.eu/research/conferences/2004/sciprof/pdf/final_en.pdf,* retrieved 10 April 2012.

European Commission (ed.) (2006), Science Education now: A Renewed Pedagogy for the Future of Europe, European Commission Directorate-General for Research, Brussels, *http://ec.europa.eu/research/science-society/document_library/pdf_06/report-rocard-on-science-education_en.pdf,* retrieved 10 April 2012.

Faubert, V. (2009), *School Evaluation: Current Practices in OECD Countries and a Literature Review*, OECD Education Working Papers, Paris.

Fensham, P.J. (2009), "Real world contexts in PISA science: Implications for context- based science education", *Journal of Research in Science Teaching,* 46(8), pp. 884-896.

Fisher, C. B., S. A. Wallace, and R. E. Fenton (2000), "Discrimination Distress during Adolescence", *Journal of Youth and Adolescence*, Vol. 29, pp. 679–695.

Fivush, R. and N.R. Hamond (1990), *Knowing and Remembering in Young Children*, University Press, New York, Cambridge.

Fleener, M. J. (1996), "Scientific World Building on the Edge of Chaos: High School Students' Beliefs about Mathematics and Science", School Science and Mathematics, 96(6), pp. 312–320.

Furtak, E.M, Seidel, T., Iverson, H., Briggs, D. (2012). Experimental and Quasi-Experimental Studies of Inquiry-Based Science Teaching: A Meta-Analysis. *Review of Educational Research, 82*(3), 300-329.

Gamoran, A., A.C. Porter, J. Smithson and P.A. White (1997)," Upgrading High School Mathematics Instruction: Improving Learning Opportunities for Low-Achieving, Low-Income Youth", *Educational Evaluation and Policy Analysis,* 19, pp. 325-338.

Gándara, P., R. Rumberger, J. Maxwell-Jolly and R. Callahan (2003), English Learners in California Schools: Unequal Resources, Unequal Outcomes, Education Policy Analysis Archives, 11(36), *http://epaa.asu.edu/epaa/v11n36*/, retrieved 03 April 2012.

Ghuman, S. and C. Lloyd (2010), "Teacher Absence as a Factor in Gender Inequalities in Access to Primary Schooling in Rural Pakistan", *Comparative Education Review*, Vol. 54, No. 4, pp. 539-554.

Gillies, J. and J.J. Quijada (2008), *Opportunity to Learn: A High Impact Strategy for Improving Educational Outcomes in Developing Countries*, Working Paper, Washington, DC.

Goodman R. (1997), “The Strengths and Difficulties Questionnaire: A Research Note”, *Journal of Child Psychology and Psychiatry*, 38, pp. 581-586.

Gottfried, A.E., G.A. Marcoulides, A.W. Gottfried and P.H. Oliver (2009), “A Latent Curve Model of Parental Motivational Practices and Developmental Decline in Math and Science Academic Intrinsic Motivation”, *Journal of Educational Psychology,* 101(3), pp. 729-739.

Greene, J.P. and M.A. Winters (2009), “The Effects of Exemptions to Florida's Test-Based Promotion Policy. Who is retained? Who Benefits Academically?”, *Economics of Education Review*, Vol. 28, No. 1, pp. 135–142.

Griffith, C A., J.W. Lloyd, K.L. Lane and M. Tankersley (2009), “Grade Retention of Students During Grades K–8 Predicts Reading Achievement and Progress During Secondary Schooling”, *Reading & Writing Quarterly*, 26(1), pp. 51–66.

Guskey, T.R. (2007), “Multiple Sources of Evidence: An Analysis of Stakeholders' Perceptions of Various indicators of student learning”, Educational Measurement, *Issues and Practice*, Vol. 26, pp. 19-28.

Gustafsson, J.E. (2008), “Effects of International Comparative Studies on Educational Quality on the Quality of Educational Research”, *European Educational Research Journal*, 7(1), pp. 1-17.

Hanushek, E. A. and L. Wößmann (2010), *The High Cost of Low Educational Performance: The Long-Run Economic Impact of Improving PISA Outcomes.* – Paris: OECD.

Hanushek, E.A. and L. Wößmann (2011), “The Economics of International Differences in Educational Achievement”, in E.A. Hanushek, S. Machin and L. Wößmann (eds.), *Handbook of the Economics of Education*, Vol. 3, North Holland, Amsterdam, pp. 89–200.

Harlen, W. and R.D. Crick (2002), *A Systematic Review of the Impact of Summative Assessment and Tests on Students' Motivation for Learning, Review conducted by the Assessment and Learning Research Synthesis Group*.

Hattie, J. A. C. (2009): *Visible Learning. A synthesis of over 800 meta-analyses relating to achievement.* London & New York: Routledge.

Hattie, J. and H. Timperley (2007), “The Power of Feedback”, *Review of Educational Research*, Vol. 77, No. 1, pp. 81–112.

Heckman, J. (2006), “Skill Formation and the Economics of Investing in Disadvantaged Children”, Science, 312, pp. 1900-1902.

Heckman J.J., J. Stixrud and S.Urzua (2006), "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior", *Journal of Labor Economics,* University of Chicago Press, Vol. 24, 3, pp. 411-482.

Hill, H.C., B. Rowan, D. Loewenberg Ball (2005), “Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement”, *American Educational Research Journal*, Jg. 42, H. 2, pp. 371–406.

Hofer, B.K. and P.R. Pintrich (2002*), Personal Epistemology: The Psychology of Beliefs about Knowledge and Knowing*, B.K. Hofer and P.R. Pintrich (eds.), Lawrence Erlbaum Associates Publishers, Mahwah, NJ, *http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2001-18187-000&site=ehost-live*, retrieved 10 April 2012.

Hord, S.M. (1997), Professional Learning Communities: Communities of Continuous Inquiry and Improvement, Southwest Educational Development Laboratory, Austin TX.

Ingersoll, R.M. and D. Perda (2010), “Is the Supply of Mathematics and Science Teachers Sufficient?”, *American Educational Research Journal*, Vol. 47, No. 3, pp. 563-694.

Jeanpierre, B., Oberhauser, K., & Freeman, C. (2005). Characteristics of professional development that effect change in secondary science teachers' classroom practices. *Journal of Research in Science Teaching*, 42(6), 668-690.

Jiménez-Aleixandre, M.P. and S. Erduran (2007), “Argumentation in Science Education: Perspectives from Classroom-Based Research”, in S. Erduran and M.P. Jiménez-Aleixandre (eds.), *Argumentation in Science Education: An Overview*, Springer, Dordrecht, pp. 3–29.

Kahneman, D., A. B. Krueger, D. A. Schkade, N. Schwarz and A. A. Stone (2004), “A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method”, *Science*, 306, pp. 1776-1760.

King, D.and M.R. Stephen (2012),” Learning Science Through Real-World Contexts”, in BJ. Fraser, K. Tobin and C. McRobbie (eds.), *Springer International Handbooks of Education, Vol. 24, Second International Handbook of Science Education*, Springer, Dordrecht, pp. 69–79.

King, G. and J. Wand (2007), “Comparing Incomparable Survey Responses: New Tools for Anchoring Vignettes”, *Political Analysis*, 15, pp. 46-66.

Klieme, E., Ch. Pauli and K. Reusser (2009). The Pythagoras Study - Investigating effects of teaching and learning in Swiss and German mathematics classrooms. In T. Janik and T. Seidel (Eds.), *The Power of Video Studies in Investigating Teaching and Learning in the Classroom*. (pp. 137-160). Münster: Waxmann

Kloosterman, R. and P.M. de Graaf (2010), “Non-Promotion or Enrolment in a Lower Track? The Influence of Social Background on Choices in Secondary Education for Three Cohorts of Dutch Pupils, *Oxford Review of Education*, Vol. 36, No. 3, pp. 363–384.

Kobarg, M., M. Prenzel, T. Seidel, M. Walker, B. McCrae, J. Cresswell and J. Wittwer (2011), *An International Comparison of Science Teaching and Learning: Further Results from PISA 2006*, Waxmann, Münster.

Koeppen, K. et al. (2008), “Current Issues in Competence Modeling and Assessment”, *Journal of Psychology*, Vol. 216, No. 2, pp. 61–73.

Krapp, A. and M. Prenzel (2011), “Research on Interest in Science: Theories, Methods, and Findings”, *International Journal of Science Education*, 33(1), pp. 27-50.

Kristjansson, B., Petticrew, M., MacDonald, B., Krasevec, J., Janzen, L., Greenhalgh, T., et al. (2009). School feeding for improving the physical and psychosocial health of disadvantaged students. *Cochrane Collaboration, 2009(3).* Retrieved January 27, 2010, from The Cochrane Library Database.

Kunter, M. (2005), Multiple Ziele im Mathematikunterricht, Pädagogische Psychologie und Entwicklungspsychologie, Band 51, Waxmann, Münster.

Kunter, M., A. Frenzel, G. Nagy, J. Baumert and R. Pekrun (2011), “Teacher Enthusiasm: Dimensionality and Context Specificity”, *Contemporary Educational Psychology*, Vol. 36, No. 4, pp. 289–301.

Lindqvist, E. and R. Vestman (2011), “The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish enlistment”, *American Economical Journal: Applied Economics*, 3(1), pp. 101-128.

Logan, J., S. Piasta, L. Justice, C. Schatschneider and S. Petrill (2011), “Children’s Attendance Rates and Quality of Teacher-Child Interactions in At-Risk Preschool Classrooms: Contribution to Children’s Expressive Language Growth”, Child and Youth Care Forum, 40 (6), pp. 457-477.

Lomos, C., R. H. Hofman, and R. J. Bosker (2011), Professional co mmunities and student achievement – a meta-analysis. *School Effectiveness and School Improvement*, 22(2), pp. 121-148.

MacKay, R. (2009), "Remarks on the Inefficient Use of Time in the Teaching and Learning of Mathematics in five Secondary Schools", in J.H. Meyer and A. van Biljon (eds.), Proceedings of the 15th Annual Congress of the Association for Mathematics Education of South Africa (AMESA) "Mathematical Knowledge for Teaching", pp. 79-85, www.amesa.org.za/AMESA2009/Volume109.pdf#page=85, retrieved 03 April 2012.

Mahoney, J.L. and R.B. Cairns (1997), "Do Extracurricular Activities Protect Against Early School Drop-Out?" *Developmental Psychology,* 33, pp. 241-352.

Markowitsch, H.J. and H. Welzer (2009), *The Development of Autobiographical Memory*, Psychology Press. London.

Martin, M.O., I.V.S. Mullis and P. Foy, P. (2008), *TIMSS 2007 International Science Report: Findings from IEA's Trends in International Mathematics and Science Study at the Eighth and Fourth Grades*, Boston College, Chestnut Hill, MA.

McHale, S.M., A.C. Crouter and C.J. Tucker (2001), "Free-Time Activities in Middle Childhood: Links with Adjustment in Early Adolescence", Child Development, 72, pp. 1764-1778.

McMillan, J.H. (2001), "Secondary Teachers' Classroom Assessment and Grading Practices", *Educational Measurement*, 20, pp. 20-32.

McMillan, J.H. (ed.) (2007), *Formative Classroom Assessment: Theory into Practice*, Teachers College Columbia University, New York.

Millot, B. and J. Lane (2002), "The Efficient Use of Time in Education", *Education Economics*, 10(2), pp. 209-228.

Minner, D.D., A.J. Levy and J. Century (2010), "Inquiry-Based Science Instruction - What is it and does it Matter?" Results from a research synthesis 1984 to 2002, *Journal of Research in Science Teaching*, 47(4), pp. 474–496.Nevo, D. (ed.) (2002), *School-Based Evaluation. An International Perspective*, JAI Press, Amsterdam [u.a.].

Neyer, F.J., J. Felber and C. Gebhardt (2012), „Entwicklung und Validierung einer Kurzskala zur Erfassung von Technikbereitschaft", *Diagnostica, 58*(2), pp. 87–99.

OECD, (2004), Learning for Tomorrow's World: First Results from PISA 2003, OECD Publishing, Paris.

OECD (2005), *PISA 2003 Technical Report*, OECD, Paris.

OECD (2005a), Formative Assessment: Improving Learning in Secondary Education, OECD, Paris, p. 21.

OECD (2005b), *The Definition and Selection of Key Competencies, Executive Summary, www.oecd.org/dataoecd/47/61/35070367.pdf,* retrieved 02 March 2012, p. 11.

OECD (2007), *Science Competencies for Tomorrow's World*, Executive Summary, OECD Publishing, Paris.

OECD (2008), *Encouraging Student Interest in Science and Technology Studies*, OECD Publishing, Paris.

OECD (2010), *PISA 2009 Results: What Students can do? Student Performance in Reading, Mathematics and Science*, Vol. 1, OECD Publication, Paris.

OECD (2011), *Quality Time for Students: Learning in and out of School,* OECD Publishing, Paris.

OECD (2011a), *How's life? Measuring Well-Being*, OECD Publishing, Paris, *http://dx.doi.org/10.1787/9789264121164-en,* retrieved 18 July 2012.

OECD (2013, *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris: OECD Publishing. Retrieved from: http://dx.doi.org/10.1787/9789264190511-en.

Ou, S.-R. and A.J. Reynolds (2010), "Grade Retention, Postsecondary Education, and Public Aid Receipt", *Educational Evaluation and Policy Analysis*, Vol. 32, No. 1, pp. 118–139.

Osborne, J. (2012). "The Role of Argument: Learning how to Learn in School Science", in B.J. Fraser, K. Tobin and C. McRobbie (eds.), *Second International Handbook of Science Education*, *Springer International Handbooks of Education*, Vol. 24, Springer, Dordrecht, pp. 933-949.

Paine, L. and L. Ma (1993), "Teachers Working Together: A Dialogue on Organisational and Cultural Perspectives of Chinese Teachers", *International Journal of Educational Research*, Vol. 19 No. 8, pp. 675-697.

Parker, P.D., H.W. Marsh, J. Ciarrochi, S. Marshall and A.S. Abduljabbar (2013), *Juxtaposing math self-efficacy and self-concept as predictors of longterm achievement outcomes*, Educational Psychology. DOI:10.1080/01443410.2013.797339

Partnership for 21st Century Skills (2008), 21st Century Skills, Education & Competitiveness: A Resource and Policy Guide.

Patall, E.A., H. Cooper and A.B. Allen (2010), "Extending the School Day or School Year", *Review of Educational Research*, 80(3), pp. 401-436.

Patrick, H., J.C. Turner, D.K. Meyer and C. Midgley (2003), "How Teachers Establish Psychological Environments during the First Days of School: Associations with Avoidance in Mathematics", Teachers College Record, 105, pp. 1521–1558.

Pellegrino, J., N. Chudowsky and R. Glaser (eds.) (2001), *Knowing what Students Know: The Science and Design of Educational Assessment*, National Academy Press, Washington, D.C.

Poropat, A.E. (2009), "A Meta-Analysis of the Five Factor Model of Personality and Academic Performance", *Psychological Bulletin*, 135(2), pp. 322-338.

Porterfield, S. and A.E. Winkler (2007), "Teen Time Use and Parental Education: Evidence from the CPS, MTF, and ATUS", *Monthly Labor Review*, May 2007, pp. 37-56.

Portner, H. (ed.) (2005), *Teacher Mentoring and Induction: The State of the Art and Beyond*, Corwin, Thousand Oaks, CA.

Prenzel, M, A. Kristen, P. Dengler, R. Ettle, and T. Beer (1996), Selbstbestimmt motiviertes und interessiertes Lernen in der kaufmännischen Erstausbildung, *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 13, pp. 108-127

Prenzel, M., T. Seidel and M. Kobarg (2012)," Science Teaching and Learning: An International Comparative Perspective", in B.J. Fraser, K. Tobin and C. McRobbie (eds.), *Second International Handbook of Science Education*, Springer International Handbooks of Education, Vol. 24, Springer, Dordrecht, pp. 667–678.

Purves A.C. (1987), I.E.A. an Agenda for the Future, *International Review of Education*, Vol. 33 (1. 1987), p. 103-107.

Rakoczy, K., E. Klieme and C. Pauli (2008), „Die Bedeutung der wahrgenommenen Unterstützung motivationsrelevanter Bedürfnisse und des Alltagsbezugs im Mathematikunterricht für die selbstbestimmte Motivation", *Zeitschrift für Pädagogische Psychologie/ German Journal of Educational Psychology*, 22(1), pp. 25-35.

Richardson, M., C. Abraham and R. Bond (2012), “Psychological Correlates of University Students’ Academic Performance: A Systematic Review and Meta-Analysis”, *Psychological Bulletin,* 138, pp. 353-387.

Roberts, B.W., N.R. Kuncel, R. Shiner, A. Caspi and L.R. Goldberg (2007), “The Power of Personality: The Comparative Validity of Personality Traits, Socioeconomic Status, and Cognitive Ability for Predicting Important Life Outcomes”, Perspectives on Psychological Science, 2, pp. 313-345.

Rowan, B., E. Camburn and R. Correnti (2004), “Using Teacher Logs to Measure the Enacted Curriculum: A Study of Literacy Teaching in Third-Grade Classrooms”, *Elementary School Journal,* 205, pp. 75-101.

Ryan, R.M. and E.L. Deci (2000), “Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions”, Contemporary Educational Psychology, 25(1), pp. 54–67.

Ryan, R.M./Deci, E.L. (2002), An overview of Self-Determination Theory: An organismic-dialectical perspective. In: Deci, E.L./Ryan, R.M. (Eds.): *Handbook of Self-Determination Research*. Rochester: University Press, S. 3-33.

Ryan, K.E., M. Chandler and M. Samuels (2007), “What Should School-Based Evaluation Look Like?”, *Studies In Educational Evaluation*, Vol. 33, 3-4, pp. 197–212.

Salmon, J., Ball, K., Crawford, D., Booth, M., Telford, A., & Hume, C. (2005), Reducing sedentary behaviour and increasing physical activity among 10-year-old children: Overview and process evaluation of the „Switch-Play“ intervention. *Health Promotion International, 20(1),* 7–17.

Sammons, P., Y. Anders, K. Sylva, E. Melhuish I. Siraj-Blatchford, B. Taggart and S. Barreau (2008), “Children’s Cognitive Attainment and Progress in English Primary Schools during Key Stage 2”, Investigating the Potential Continuing Influences of Pre-School Education, *Zeitschrift für Erziehungswissenschaft*, Sonderheft 11, pp. 179-198.

Sanders, James R.; Davidson, E. Jane (2003): A Model for School Evaluation. In: Thomas Kellaghan, Daniel L. Stufflebeam und Lori A. Wingate (Hg.): *International handbook of educational evaluation.* Dordrecht: Kluwer Acad. Publ. (Kluwer international handbooks of education, 9), S. 807-826.

Santiago, P. and F. Benavides (2009), *Teacher Evaluation: A Conceptual Framework and Examples of Country Practices*, Paris.

Scheerens, J. and R.J. Bosker (1997). *The Foundations of Educational Effectiveness*, Pergamon Press, Oxford.

Scheerens, Jaap (2002): School self-evaluation. Origins, definition, approaches, methods and implementation. In: David Nevo (Hg.): *School-based evaluation. An international perspective*. Amsterdam [u.a.]: JAI Press, S. 35-69.

Scheerens, J., C.A.W. Glas and S. Thomas (eds.) (2003), *Educational Evaluation, Assessment, and Monitoring. A Systemic Approach,* Contexts of Learning, Swets & Zeitlinger, Lisse Netherlands, Exton, PA.

Scherff, Lisa, and C.L. Piazza (2008), “Why Now More Than Ever, We Need to Talk about Opportunity to Learn”, *Journal of Adolescent and Adult Literacy,* 52, No. 4, pp. 343-352.

Schleicher, A., (ed.) (2012), *Preparing Teachers and Developing School Leaders for the 21st Century: Lessons from around the World*, OECD Publishing, Paris.

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.

Schmidt, W.H. and A. Maier (2009), “Opportunity to Learn”, Chapter 44, in G. Sykes, B. Schneider and D.N. Plank (eds.), *Handbook of Education Policy Research*, Routledge, New York ,pp. 541-559.

Schmidt, W.H., C.C. McKnight, L.S. Cogan, P.M. Jakwerth and R.T. Houang (1999), *Facing the Consequences: Using TIMSS for a Closer Look at US Mathematics and Science Education*, Kluwer Dordrecht, NL.

Schmidt, W.H., H.C. Wang and C.C. McKnight (2005), Curriculum Coherence: An Examination of US Mathematics and Science Content Standards from an International Perspective, *Journal of Curriculum Studies*, Vol. 37, Issue 5, pp. 525-559.

Schmidt, W.H. and 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*, Chapter 4, Brookings Institute Press, Washington, DC.

Scriven, M. (2003), “Evaluation Theory and Metatheory. in T. Kellaghan, D.L. Stufflebeam and L.A. Wingate (eds.), *International Handbook of Educational Evaluation,* Kluwer International Handbooks of Education, No. 9, Kluwer Acad. Publ., Dordrecht, pp. 15–31.

Seidel, T. and R.J. Shavelson (2007), “Teaching Effectiveness Research in the Past Decade: The Role of Theory and Research Design in Disentangling Meta-Analysis Results*”, Review of Educational Research,* 77(4), pp. 454-499.

Shayer, M. & Adhami, M. (2007). “Fostering Cognitive Development Through the Context of Mathematics: Results of the CAME Project”, *Educational Studies in Mathematics,* 64 (3), pp. 265 – 291.

Shepard, L.A. (2006), “Classroom Assessment”, in R.L. Brennan (ed.), Educational Measurement, Praeger Publishers, Westprot, CT, pp. 623–646.

Shulman, L. S. (1985), “Paradigms and Research Programs in the Study of Teaching: A contemporary perspective”, in M C. Wittrock (ed.), *Handbook of Research on Teaching* (3rd ed, pp. 3–36), Macmillan, New York.

Shute, V.J. (2008), “Focus on Formative Feedback”, *Review of Educational Research*, Vol. 78, No. 1, pp. 153–189.

Spurrier, N., Sawyer, M., Clark,J., & Baghurst, P. (2003). Socio-economic differentials in the health-related quality of life of Australian children: Results of a national study. *Australian and New Zealand Journal of Public Health, 27(1),* 27–33.

Stanat, P. and G. Christensen (2006), *Where Immigrant Students Succeed -A Comparative Review of Performance and Engagement in PISA 2003*, OECD, Paris.

Steinert, B., E. Klieme, K. Maag Merki, P. Döbrich,U. Halbheer and A. Kunz (2006), “Lehrerkooperation in der Schule“ (Teacher co-operation in schools), *Zeitschrift für Pädagogik,* 52 ;2 , pp. 185-203.

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.

Stiggins, R.J. (2007), “Conquering the formative assessment frontier”, in J.H. McMillan (ed.), *Formative classroom assessment. Theory into practice,* Teachers College Columbia University, New York, pp. 8–28.

Supovitz, J.A. and H.M. Turner (2000), “The Effects of Professional Development on Science Teaching Practices and Classroom Culture”, Journal of Research in Science Teaching, Vol. 37 No. 9, pp. 963-980.

Sykes, G. (1996), “Reform of and as Professional Development”, Phi Delta Kappan, Vol. 77, pp. 465–467.

Sykes G., B. Schneider and D.N. Plank (eds.) (2009), *Handbook of Education Policy Research*, Routledge, New York.

Sylva, K., E. Melhuish, P. Sammons, I. Siraj-Blatchford, and B. Taggart, (2011), Pre-School Quality and Educational Outcomes at Age 11: Low Quality has Little Benefit, *Journal of Early Childhood Research*, 9 (2), pp. 109-124.

Sylva, K., A. Stein, P. Leach, J. Barnes, L.-E. Malmberg, and the FCCC Team (2011), Effects of Early Child-Care on Cognition, Language, and Task-Related Behaviours at 18 months: An English Study, *British Journal of Developmental Psychology*, 29 (1), pp. 18-45.

Taylor, J.A., M.A.M. Stuhlsatz and R.W. Bybee (2009), Windows into High-Achieving Science Classrooms, in R.W. Bybee and B. McCrae (eds.), P*ISA Science 2006, Implications for Science Teachers and Teaching*, NSTA Press, Arlington, pp. 123–13.

The PROBE Team (1999), *Public Report on Basic Education in India*, Oxford University Press, New Delhi.

Timperley, H. et al. (2007), *Teacher Professional Learning and Development: Best Evidence Synthesis Iteration [BES]*, Wellington, Ministry of Education, New Zealand, www.educationcounts.govt.nz/goto/BES, retrieved 15 April 2011.

Travers, K.J. and I. Westbury (1989), *The IEA Study of Mathematics I: Analysis of Mathematics Curricula*, Pergamon Press, Oxford.

Tschannen-Moran, M. and A. Woolfolk Hoy (2001), "Teacher Efficacy: Capturing an Elusive Construct", *Teaching and Teacher Education*, 17 (7), pp. 783-805.

Turner, J.C., D.K. Meyer, K.E. Cox, C. Logan, M. DiCintio and C.T. Thomas, (1998), "Creating Contexts for Involvement in Mathematics", Journal of Educational Psychology, 90, pp. 730–745.

UNESCO (2006), *Education for all Global Monitoring Report 2007: Strong Foundations*, UNESCO, Paris.

Vieluf, S., D. Kaplan, E. Klieme and S. Bayer (in press), Profiles of Teaching Practices and Insights into Innovation: Results from TALIS 2008, OECD, Paris.

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.

Van de Vijver, F. J. R., S. M. Breugelmanns, and S. R. G. Schalk-Soekar (2008), "Multiculturalism: Construct validity and stability", International Journal of Intercultural Relations, Vol. 32 No. 2, pp. 93–104.

Valenti, J.E. and D.H. Tracey (2009), "Full-day, half-day, and no Preschool Effects on Urban Children's First-Grade Reading Achievement", Education and Urban Society, 41 (6), pp. 695-711.

Wang, L., C. MacCann, X. Zhuang, O. Lydia Liu and D.R. Roberts (2009), „Assessing teamwork and collaboration in high school students: A multimethod approach", *Canadian Journal of School Psychology*, 24(2), 108-124.

Wigfield, A., J.S. Eccles and D. Rodriguez (1998), "The Development of Children's Motivation in School Contexts", Review of Research in Education, 23, p. 73.

Willms J.D. (2006*), Learning Divides: Ten Policy Questions about the Performance and Equity of Schools and Schooling Systems*, UNESCO Institute for Statistics , Montreal, Canada.

Wilson, M. (ed.) (2004), *Towards coherence between classroom assessment and accountability*, Yearbook of the National Society for the Study of Education, University of Chicago Press, Chicago, Ill.

Walton, G.M. & Cohen, G.L. (2011). A Brief Social-Belonging Intervention Improves Academic and Health Outcomes of Minority Students. *Scienec, Vol. 331 No. 6023*, pp. 1447-1451.

Wong, C. A., J. S. Eccles, and A. Sameroff (2003), “The Influence of Ethnic Discrimination and Ethnic Identification on African American Adolescents’ School and Socio-emotional Adjustment”, *Journal of Personality*, Vol. 71 No. 6, pp. 1197-1232.

World Health Organization (2008b). *School Policy framework: Implementation of the WHO global strategy on diet, physical activity and health.* WHO Press, Geneva.

Wößmann, L., E. Lüdemann, G. Schütz and M.R. West (2007), “*School Accountability, Autonomy, Choice, and the Level of Student Achievement: International Evidence from PISA 2003”*, OECD Education Working Paper No. 13.

Wößmann, L., E. Luedemann, G. Schuetz and M.R. West (2009), School Accountability, Autonomy and Choice around the World, Edward Elgar Publishing, Massachusetts.

Appendix 1

Indicators using information from PISA Questionnaires in OECD's Education at a Glance, 2002-2012

Policy Issue	Indicators (Year, Number)	PISA Questionnaire Variables Used
A The Output of Educational Institutions and the Impact of Learning		
Impact of student and family background on cognitive performance	2012-A5/A6, 2011- A5/A6, 2009-A4, 2008-A5, 2007-A6, 2002-A7, 2006-A6/A7, 2004-A9, 2003-A8/A11, 2002-A7/A9/A10	Gender, immigrant status, language spoken at home, mother's educational level, SES (HISEI), career expectations, instrumental motivation, self-regulated learning
Non-cognitive outcomes	2007-A4/A5, 2004-A8 2003-A9/A10	Career expectations, Attitudes towards mathematics, approaches to learning, Sense of Belonging (well-being at school), Reading engagement, Indices of self-regulated learning
Relations of cognitive and non-cognitive outcomes	2012-A4, 2009-A5 2011-A6	Enjoyment of science, science-related activities, regular lessons and out-of-school lessons in science, instrumental motivation, future-oriented motivation to learn science, importance of doing well, science-related careers (school preparation, student information), career expectations; Enjoyment of reading, time spent on reading, diversity of reading material; Student-level input: gender, SES, immigrant status
Parental perceptions	2008-A6	Parental SES, Child's past performance, perceived school standards, disciplinary atmosphere, general evaluation of the school, teacher competence and dedication, instructional content and measures, monitoring and reporting
C Access to Education, Participation, and Progression		
Student performance by type of programme	2008-C1, 2007-C1	Programme orientation (academic vs. vocational)
D The Learning Environment and Organisation of Schools		
School type	2005-D5	Type of school (public vs private)
School climate	2002-D5	Principals' perception of factors affecting school climate, teacher morale, school autonomy, teacher-student-relations, disciplinary climate, achievement pressure, homework policy, quality and use of school resources
Equity in outcomes and opportunities	2011-D6	SES, gender, immigrant status, parental education, value for schooling outcomes, student-teacher relations, disciplinary climate
Access to and use of ICT	2006-D5, 2002-D3/D4	ICT resources and computers at school, shortage of ICT resources, perceived comfort and availability in using ICT, frequency of using computers at different places, gender

Appendix 2

Selected analytical models used in publications on the PISA 2006 data for contexts of science achievement

Publication	Research Question or Model
Drechsel, Carstensen, Prenzel (2011)	Dimensionality of science interest
Olson & Lie (2011)	Country- and culture specific profiles of interest
Anley & Ainley (2011a)	Students’ enjoyment, learning engagement, and achievement
Ainley & Ainley (2011b)	Knowledge, affect, value, and students' interest in science
Lavonen & Laaksonen (2009)	Learning activities, interest in science, self-efficacy, self-concept, and performance
Fensham (2009)	Gender, task context and science performance
Buccheri, Gruber, Bruhwiler (2011)	Gender specificity in interest and vocational choices
Mc Conney et al. (2011)	Science interests among minority students
Luu & Freeman (2011)	Scientific literacy and ICT-related variables
Kubiatko & Vlckova (2010)	
Ho (2010)	Parental involvement and students' science performance.
Basl (2011)	Explaining interest in future science-related careers
Kjaernsli & Lie (2011)	
Willms (2010)	School composition, school and classroom context, and students' literacy skills
Dincer & Uysal (2010)	Effects of school programme types
Coll et al. (2010)	influence of educational context in a western vs. Asian country