

## **Symposium on Assessing Policy Lessons from PISA**

18-20 November 2002, Berlin

### **Drawing the lessons from PISA 2000 – Long-term research implications: Gaining a better understanding of the relationship between system inputs and learning outcomes by assessing instructional and learning processes as mediating factors**

Juergen Baumert (Max Planck Institute for Human Development)

Werner Blum (University of Kassel) & Michael Neubrand (University of Flensburg)

Not least because of the rapid succession of assessment cycles, PISA has developed a pace of its own that keeps all those involved in a permanent state of alertness. Most of you know what I am talking about. This is what makes occasions such as this conference all the more important. They give us the opportunity to pause for breath, take stock of what has been achieved thus far, and on this basis to identify areas in need of optimisation as well as potential for future developments.

I would like to begin by recapping on the strengths of PISA and highlighting three features in particular:

- (1) the unusually well-elaborated theoretical and methodological underpinning of PISA's core assessment programme,
- (2) the systematic extension of the assessment programme beyond the core competencies of reading, mathematical and scientific literacy, and
- (3) the combination of strictly standardised international data-collection procedures and rigorous quality standards with national research options. The national options not only ensure a better fit between the international PISA assessments and national concerns, but can also contribute to the further development of the international programme.

The findings from PISA 2000 provide a good description of the structure of core competencies, their patterns of distribution in the student population, and their relationships with social background characteristics. For most of the participating countries, this makes PISA a useful – if not indispensable – instrument for ensuring the transparency of educational

outcomes. Transparency is a necessary though not sufficient precondition for improving and optimising educational systems.

Furthermore, PISA shows what can be achieved in different countries at reasonable costs. The study furnishes compelling evidence that it is possible to optimise both the level and the distribution of the competencies acquired at school. At the same time, it provides international benchmarks.

This is a very good starting point for the PISA 2003 assessment cycle. Yet the successes of the programme also draw attention to challenges that have still to be overcome. PISA has not yet provided answers – or rather, satisfactory answers – to two important questions. The first of these questions is as follows: To what extent do the core competencies assessed in PISA serve as personal resources that shape the individual biography in terms of personal growth, career success and societal participation and – even more difficult – as societal resources contributing to the development of the national human capital? The first part of this question can only be addressed by means of a longitudinal study, beginning with a PISA assessment at the latest and focusing on how young adults cope with critical life events over the next 10 to 15 years. Most of the OECD member countries participating in PISA have not yet conducted assessments of this kind (exceptions here are the USA, Canada, Great Britain and Australia). The findings available thus far point to a cumulative effect in the sense of the “Matthew Effect” and thus lend support to the assumption that the competencies assessed in PISA are indeed key qualifications.

The second question is as follows: Assuming that the starting assumption holds, and that PISA assesses core competencies, what is it that makes one country more successful than another in preparing young people to meet the challenges of the future? Supplying answers to this question is a key aspect of the programme. PISA aims to provide the governments of the participating countries with instrumental knowledge about where and how their educational systems, and particularly the processes of teaching and learning, can be improved. This objective is one of the great strengths of PISA, and ensures a clear focus geared towards interventions on the system level. Moreover, the obligation to address the second question gives the study a particular dynamic strength. At the same time, however, it conceals a potential weakness that must be approached with due care and responsibility. The political commitment of the programme entails a serious risk that bivariate correlations between

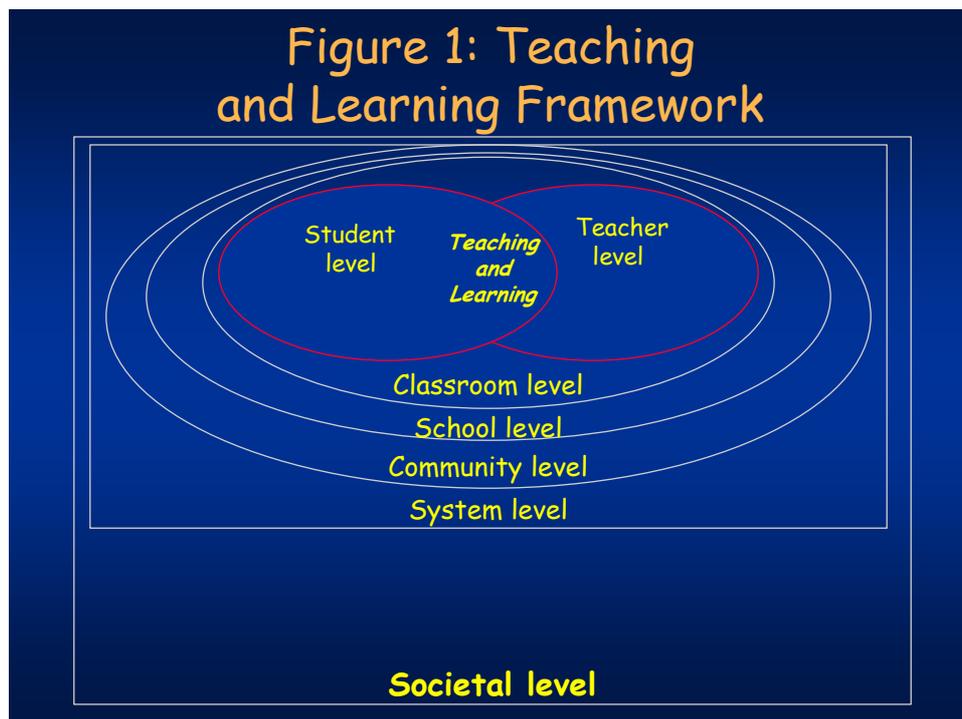
educational outcomes and properties of the system that can be influenced by political means will be interpreted as direct causal relationships.

To answer the question of what makes one country more successful than another, we have to explain cross-country differences in the relevant measurement criterion – be it the level of performance, the distribution of performance scores, or the relationship between social background and performance – by reference to factors or relations located on the system level. This is where the real problems begin. The performance of a school system is dependent on networks of interacting conditions. These are only indirectly connected to outcome variables, by way of complex mediating processes. Furthermore, most system variables are not integral characteristics of the system, but are obtained by aggregating properties located on other levels to the country level. The aggregation of most system variables involves specific problems and pitfalls that are often difficult, and sometimes impossible, to discern.

In such situations of multiple conditionality, it is common practice to perform multivariate analyses taking account of the hierarchical structure of the data. In the case of the international comparison, however, we have a small number of countries with a multitude of variables. Moreover, the less well elaborated the theoretical structure, the more variables there tend to be. This prevents the application of complex multivariate tools, and the problem is essentially insurmountable. It means that, in many cases, quantitative analyses on the level of the international comparison will be restricted to bivariate analyses or to focus on a small, theoretically well-founded selection of variables. Of course, such analyses can be of great value, especially when they are based on a well-grounded theoretical perspective. It is important, however, that one is aware of the explorative character of such analyses, and of the hypothetical nature of any conclusions drawn. Most of the correlational analyses presented in PISA 2000 are, in a manner of speaking, elements of a heuristic toolbox. The value of the tools it contains is largely dependent on the expertise of the user and on the intelligent country-specific interpretation of the study's findings.

Two implications for the future development of the PISA assessment programme emerge thus far:

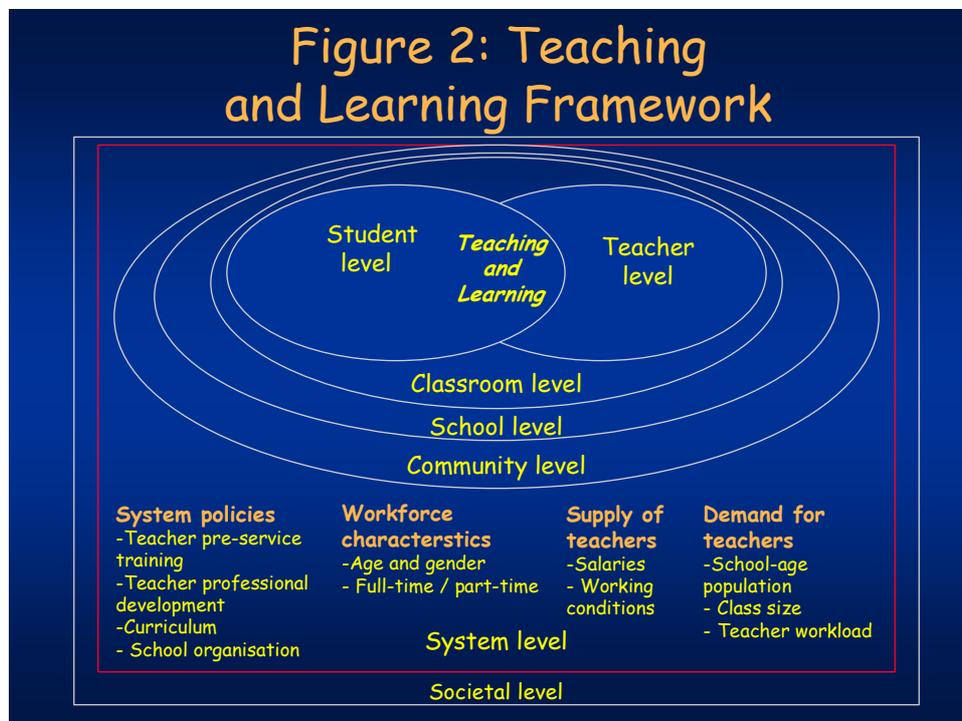
1. It is probably not wise to broaden the spectrum of system and context variables assessed in PISA unsystematically, thus adding further explorative bivariate analyses to those already existing.
2. It would seem more astute to take a theory-driven approach to filling explanatory gaps. A long-term goal here could be to reduce the explanatory gap between student performance and system or context variables that can be influenced by political means, by moving the mediating factors of successful teaching and learning processes to the centre of the analysis. After all, the performance of the system is ultimately dependent on the intelligent dovetailing of learning opportunities and the individual construction of knowledge on the part of the student.



The OECD has recently begun to pursue this goal with a draft "Conceptual Framework for Indicators of Teaching and Learning" presented by the Networks A and C Task Force on Teaching and Learning (cf. Figure 1). The framework proposed by the Task Force is based on Bronfenbrenner's ecological model of human development. According to Bronfenbrenner's model, the individual is embedded within proximal and distal layers of

environmental systems that interact with one another and with the developing person to influence development. Likewise, the draft framework distinguishes various hierarchically ordered environmental layers. The teaching and learning process is situated in the centre of the model, while system factors make up the outer layer. The Task Force regards this conceptual framework as a heuristic tool that can be used to bring structure to the field, to identify relevant factors in the interactions between the layers, and to map out theory-driven connections between micro- und macro characteristics based on the existing research literature. The aim is to use this framework to identify weaknesses in the PISA context questionnaires and to close gaps detected using theory-based, parsimonious measures.

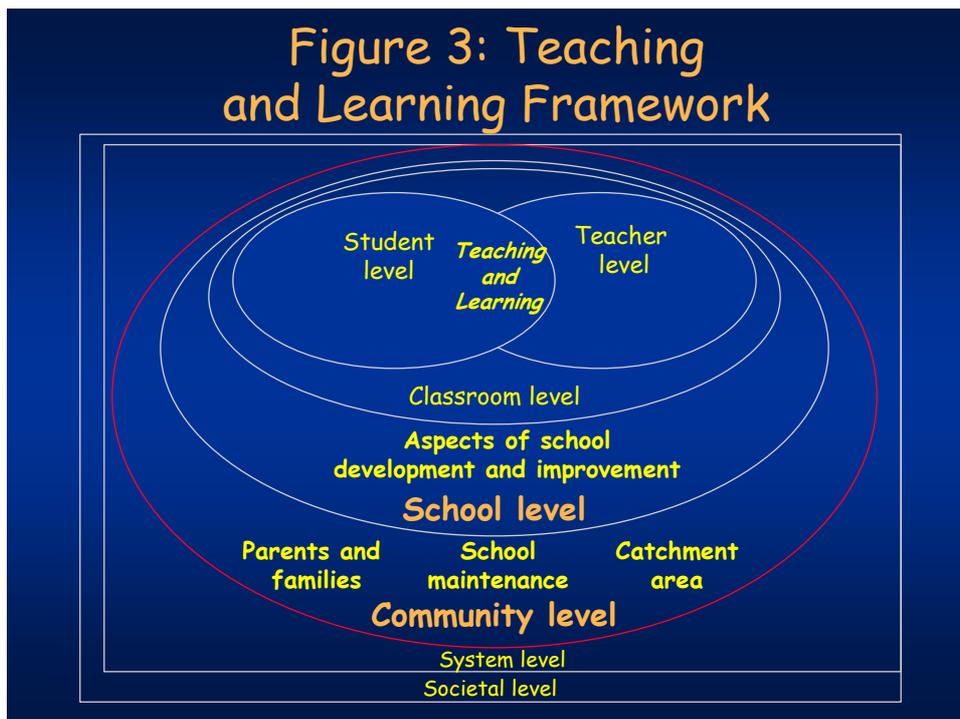
Two further layers can easily be added to the Task Force's basic model. First, schools are embedded in a community environment comprising the students' families and features of the catchment area in particular. Second, the societal level – providing a more or less supportive environment for the school system as a whole – should be added as the outermost layer (cf. Figure 1).



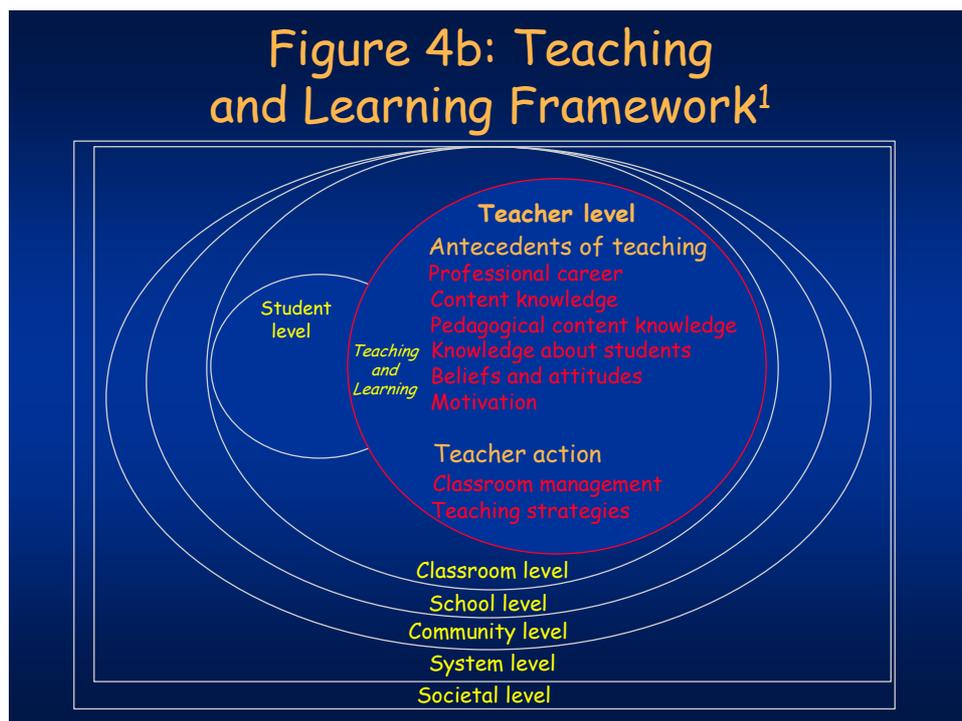
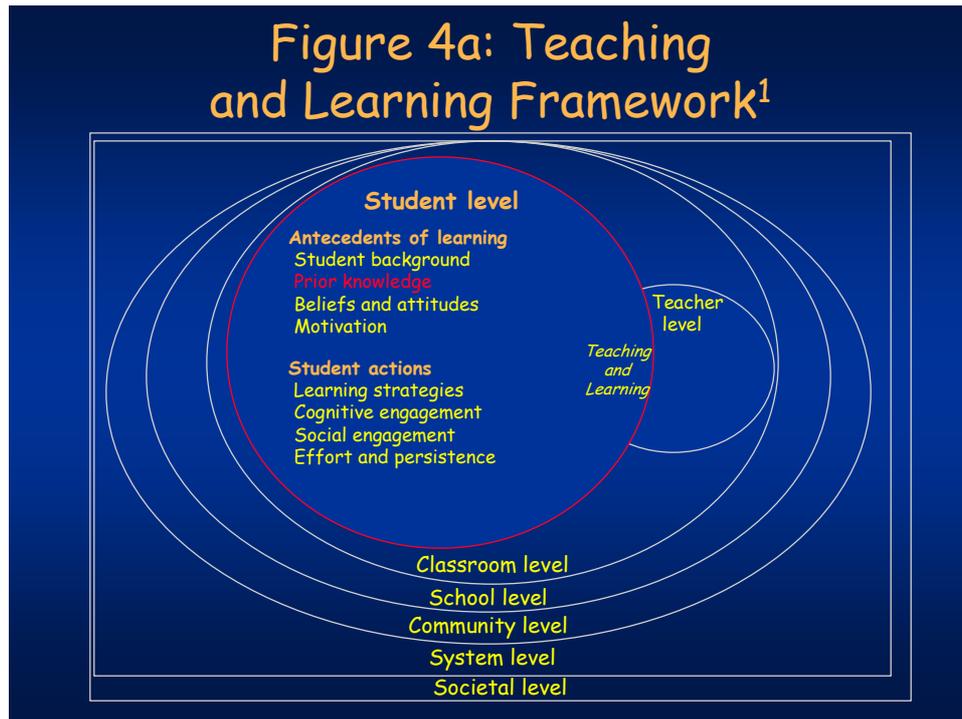
In the attempt to identify relevant factors that impact on the quality of teaching and learning

processes, the Task Force distinguishes system-level categories such as the following (cf. Figure 2):

- teaching and learning-relevant school system policies (teacher pre-service training/certification, teacher professional development, school and classroom organisation, curriculum, teacher workload, evaluation and accountability of teachers and schools)
- teacher-workforce characteristics (age, gender, ethnicity distribution, qualification and experience, full-time or part-time status)
- factors influencing the supply of teachers (salaries, career structure, working conditions, certification requirements), and
- factors influencing the demand for teachers (school-age population, class size, teacher workload).



Within the community layer, family and catchment area characteristics are the most important ones. Nested within this layer are school-level factors. All factors that have been discussed in the context of school development and school improvement come into play here. Situated within the school layer is the classroom level. Factors identified on this level are peers, social climate, and enacted curriculum (cf. Figure 3).



<sup>1</sup> Variables not sufficiently covered in PISA 2000 are shown in red.

Within the classroom layer are the teacher and student levels. According to the OECD Task Force, both of these levels are classified into actor antecedents and actions – that is, antecedents of teaching and teacher actions at the teacher level, and antecedents of learning and student actions at the student level (cf. Figures 4a and 4b). These two inner levels are

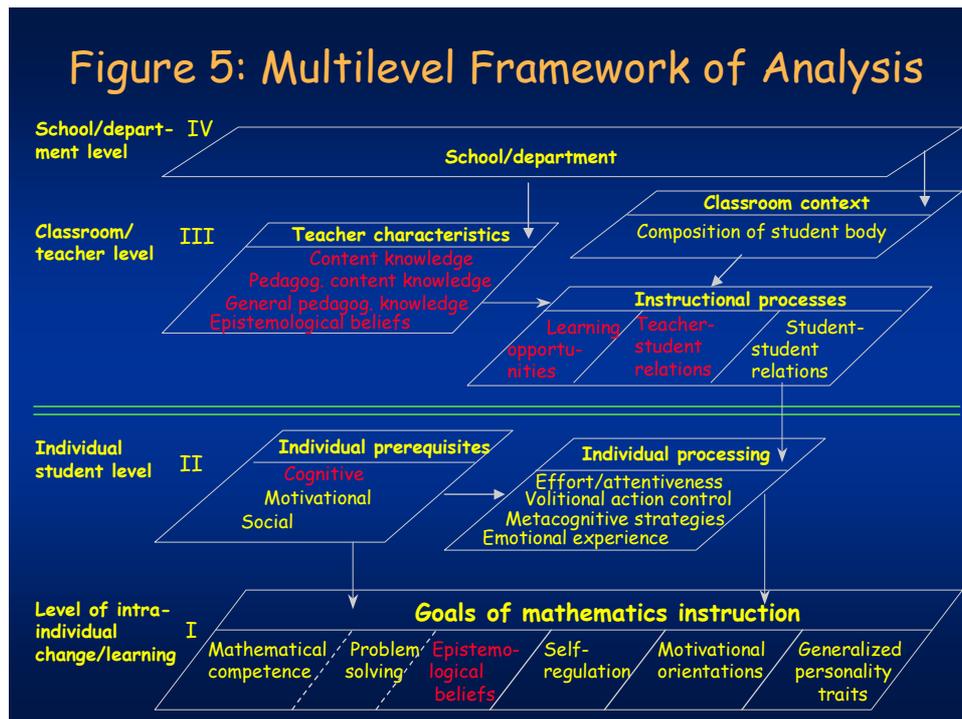
shown as having an overlapping area, represented as teaching and learning – or the interaction between teachers and students that ideally results in student growth.

All of the levels making up this framework – society, system, community, school, classroom, teacher and student – are conceptualised as having multidirectional influences. For example, just as classroom level factors can influence individual students, students help to shape what goes on in the classroom as a shared environment.

Upon closer inspection, it soon becomes clear that this teaching and learning framework is indeed a very good heuristic tool for systematically identifying gaps in the existing PISA assessment programme. Undeniably, the interaction of teaching and learning is merely implied in the existing programme; there is no real conceptual representation of this interface. With the exception of prior knowledge, the current questionnaires tap an exemplary range of student-level variables, but teacher-level variables within the classroom are not yet addressed. No account is taken of teachers' professional background, declarative knowledge, belief systems or motivation, or indeed of their procedural knowledge and professional action. At the same time, conditions of teacher action at the school and classroom level are defined on the system level and relevant data is collected, in part at least. Yet the decisive factors that mediate between student performance and system-level variables are not yet tapped. In order to be able to draw more reliable conclusions about the relevance that policy decisions at the system level have for enhancing processes of teaching and learning, we will have to work on filling these gaps in future PISA assessments.

As a feasibility study designed to test whether teacher and teaching characteristics can be examined effectively in the context of a large-scale assessment programme, the PISA 2003 cycle in Germany hopefully will include a longitudinal component focusing on the teaching and learning process in mathematics. As we have seen, the conceptual framework of the OECD Task Force seems to be a good heuristic instrument for bringing structure into the complex pattern of conditions that impact on the acquisition of competencies. It is apparent that the conceptual gaps in the existing PISA programme are found on the level of teacher characteristics and instructional processes. Therefore, the feasibility study focuses on the question of how insightful learning can be facilitated in classrooms.

As a non-recursive model, in which everything is related to everything else, however, the OECD model is less suitable as a framework for empirical analyses. The feasibility study planned in Germany does not intend to model non-recursive relationships, but instead assumes directional causal relationships. The model is presented in Figure 5. The fields of the model that are either insufficiently covered or not covered at all in the existing PISA programme are shown in red.



### Figure 6: What is Insightful Learning?

1. Insightful learning is an active process of mental construction that modifies knowledge structures.
2. Insightful learning is dependent on prior knowledge.
3. Insightful learning is cumulative; nevertheless it is always situated and mostly occurs in social situations.
4. Insightful learning is regulated by motivation and metacognitive processes (planning, control, evaluation).
5. Insightful learning is supported by mechanisms that reduce cognitive load (chunking, automatisisation).

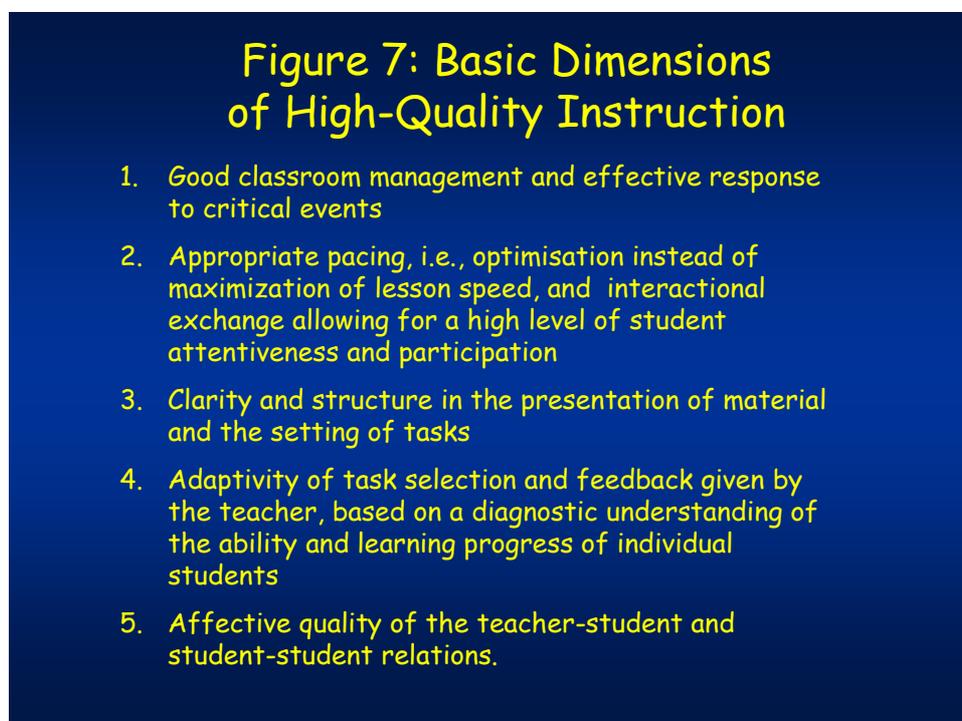
Despite certain differences on points of detail in research on instruction and learning, there is a broad consensus about some central principles of insightful learning (Baumert & Koeller 2000) (cf. Figure 6).

- Insightful learning is an active individual process of construction, in the course of which knowledge structures are modified, extended, cross-linked, hierarchically ordered or generated. Mindful receptive learning is also a constructive activity in this sense. Even in rote learning, understanding-based support strategies play a role. The decisive factor in insightful learning is active mental processing, which occurs in the active engagement with the social or natural environment or with symbolic representations. The opportunity structure of a learning environment should thus be evaluated in terms of the extent to which it supports, fosters or impedes this kind of mental activity.
- Insightful learning is dependent on individual cognitive prerequisites, and particularly on domain-specific prior knowledge. The scope and organisation of the available knowledge base determines the quality and ease of further learning. The more demanding a task or problem, the more important a student's prior knowledge becomes.
- Despite its cumulative nature, insightful learning is always situated – it inevitably occurs in a given, mostly social context. Knowledge thus contains what might be referred to as the index of the situational context of acquisition. To a certain extent, learning is always anchored in the life world – be it as artificial as the life world of the school. Because knowledge is situated in the context of acquisition, there is an inherent limitation on the transferability of knowledge acquired. To broaden the area of transfer, it is necessary to vary the contexts of acquisition and application.
- Insightful learning is also regulated by motivation and metacognitive processes of planning, monitoring, and evaluation.
- Insightful learning is supported by mechanisms that reduce cognitive load. These include chunking and the automatisisation of processes of thought and action.

Insightful learning thus evolves through the active and intelligent use of rich learning opportunities. This proactive person-environment interaction cannot be replaced by

technology. The most important aspect of the learning supply side that can be optimised is the structure and quality of learning opportunities. The horizontal green lines in Fig. 5 represent the fundamental division between the supply and usage of learning opportunities.

There is now broad consensus in empirical instructional research that there is no recipe – in the form of a single instructional conception, strategy or method – for improving instructional quality. Rather, instructional quality results from the orchestration of different strategies and methods. This process is regulated by instructional goals, the social structure of the learner group, situational conditions and not least the teacher's action repertoire.



Despite the consensus that high-quality instruction results from a flexible orchestration of strategies and methods, we should not overlook the fundamental conditions for high-quality instruction that have been identified in research on teaching and learning. These basic properties are (Helmke & Weinert 1997) (cf. Figure 7):

- good classroom management and effective response to critical events,
- appropriate pacing – i.e., optimisation as opposed to maximisation of lesson speed – and an interactional exchange allowing for a high level of student attentiveness and participation,
- clarity and structure in the presentation of material and the setting of tasks,

- adaptivity of task selection and feedback given by the teacher, based on a diagnostic understanding of the ability and learning progress of individual students, and
- the affective quality of the teacher-student and student-student relations.

What is common to all these dimensions is that they describe general characteristics of high-quality instruction, but are distal to the domain-specific learning processes themselves. They seem to be necessary, but not sufficient, conditions for cognitively activating instruction.

Recent research with more of a constructivist orientation draws attention to the significance that demanding and open-ended tasks, the structural and contextual variation of such tasks, the teacher's response to student errors, and the acquisition and automatising of cognitive and metacognitive strategies have for high-quality instruction.

These findings draw attention to the significance of teachers' domain-specific expert knowledge. The video analyses conducted in the framework of TIMSS and the TIMSS-Repeat have shown that there can be considerable cross-country differences in the intelligent choreography of domain-specific opportunities for insightful learning in the classroom.

In the context of the feasibility study, a parsimonious model of teacher expertise distinguishing different facets and types of teachers' professional knowledge was developed, as shown in Figures 8 and 9a-9c.

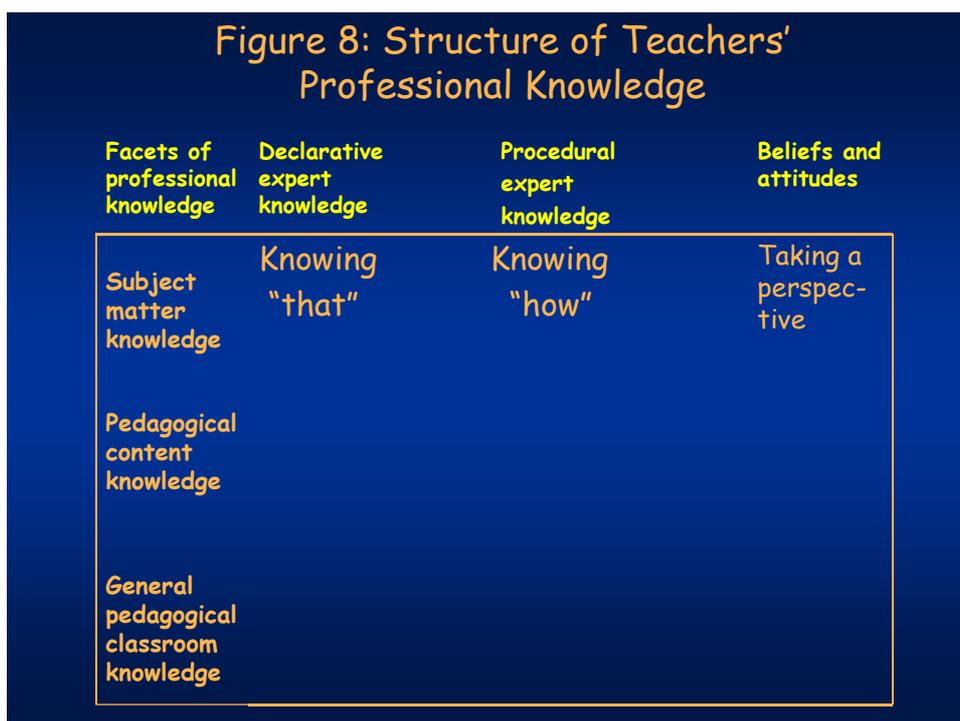
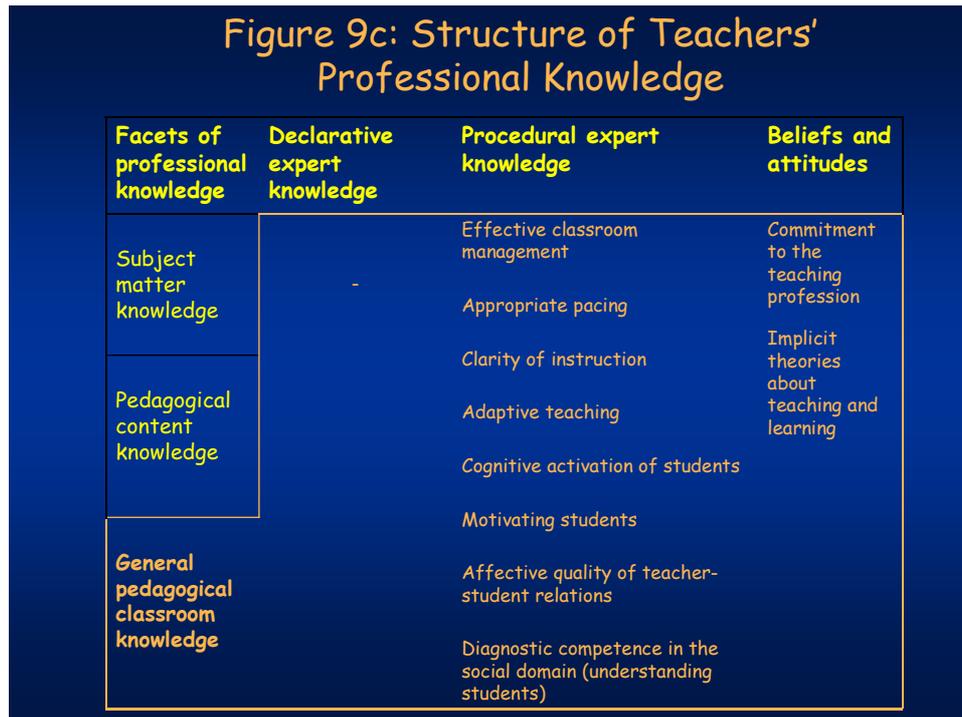


Figure 9a: Structure of Teachers' Professional Knowledge

Facets of professional knowledge	Declarative expert knowledge	Procedural expert knowledge	Beliefs and attitudes
Subject matter knowledge	Content knowledge in mathematics	-	Epistemological beliefs about mathematics
Pedagogical content knowledge	Deep understanding of basic mathematical ideas		
General pedagogical classroom knowledge			

Figure 9b: Structure of Teachers' Professional Knowledge

Facets of professional knowledge	Declarative expert knowledge	Procedural expert knowledge	Beliefs and attitudes
Subject matter knowledge	Structure of the mathematics curriculum	Selecting mathematical tasks and problems	Goals of instruction
Pedagogical content knowledge	Cognitive demands and pedagogical potentials of mathematical problems	Handling tasks in classroom instruction	Preference for specific models of teaching mathematics
	Student preconceptions	Sequencing tasks in classroom instruction	Enthusiasm about teaching mathematics
General pedagogical classroom knowledge	Typical student difficulties in understanding mathematical problems	Responding to unexpected student ideas	
		Responding to student mistakes	
		Organizing cognitively demanding practice in classroom	
		Selecting and sequencing tasks for homework and tests	
		Diagnostic competence in the field of mathematics	



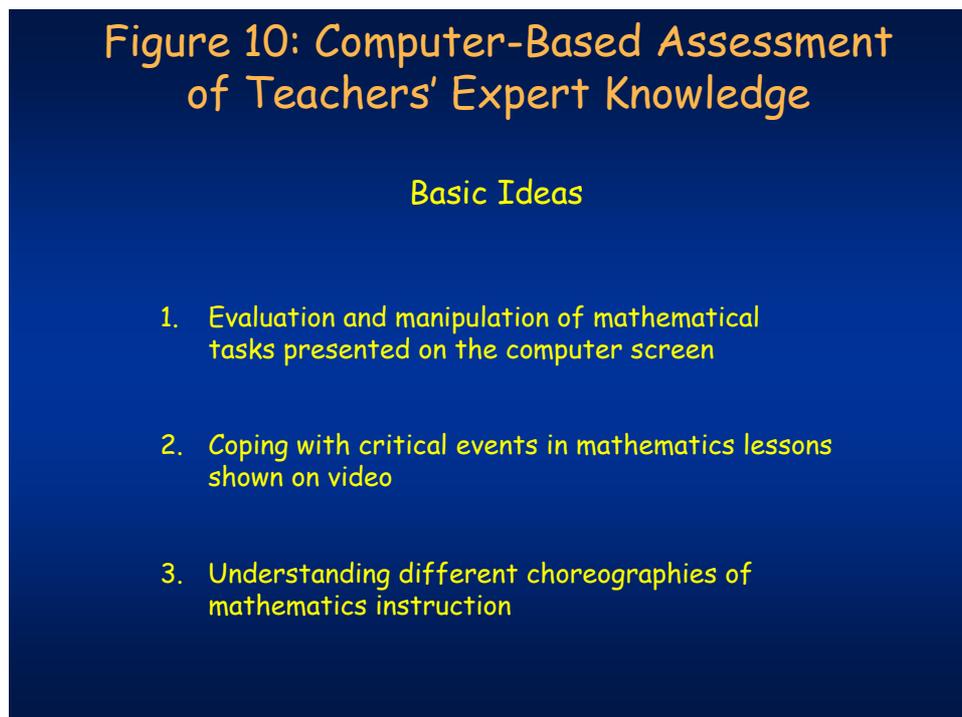
Our research focus is on what Lee Shulman (1987) termed pedagogical content knowledge and the related beliefs and attitudes. In other words, the model focuses on the expert knowledge needed to transform subject matter into insightful learning opportunities. The most important features are presented in Figure 9b.

This conceptual framework also raises the problem of data collection, however, particularly in the context of a large-scale assessment programme. To date, the results of written teacher surveys on instructional processes and practices have been disappointing, and there is no reason to believe that traditional standardised questionnaires facilitate the valid assessment of professional expertise. In the context of the TIMSS video study, which was embedded in a one-year longitudinal study in Germany, we asked students, teachers, and independent observers to rate video-taped instruction with respect to the same theoretical constructs. We ensured that the wording of the items and instructions was as uniform as possible. Nevertheless, the concurrent validity of the three rating methods was remarkably low (Clausen 2002). Teachers, students and independent observers perceive instruction from different perspectives and focus on different facets. Nevertheless, all three approaches were shown to have differential validity where the prediction of learning progress is concerned.

- Students seem to be experts on the general affective quality of instruction, the efficacy of classroom management and infrequent classroom events.

- Independent observers are distinguished by their relatively reliable evaluation of the overall quality of instruction with high inference ratings. However, their reports are also reliable when assessing frequent events with low inference ratings.
- Teacher reports on their own instruction are characterised primarily by a self-serving bias. However, their assessments do seem to become noteworthy when focusing on dimensions that can only be evaluated by raters with the necessary professional expertise and, in particular, the necessary pedagogical content knowledge. Precisely these aspects are barely ever addressed in traditional questionnaires, however.

Indeed, traditional questionnaire methods do not tend to elicit professional expertise. The goal of a profitable teacher survey must be to involve teachers in a kind of conversation about their work and to give them the opportunity to document their professional expertise. Does this imply that we should be using qualitative interviews in the context of large-scale assessment studies? Clearly, this is not a viable alternative.



Instead, we have decided to develop a computer-based assessment instrument to engage mathematics teachers in conversation about their professional activities (cf. Figure 10). This instrument, which is being evolved in a research project funded by the German Research Foundation (Baumert, Blum & Neubrand 2001), allows us:

- to present and manipulate mathematical tasks on the computer screen, and

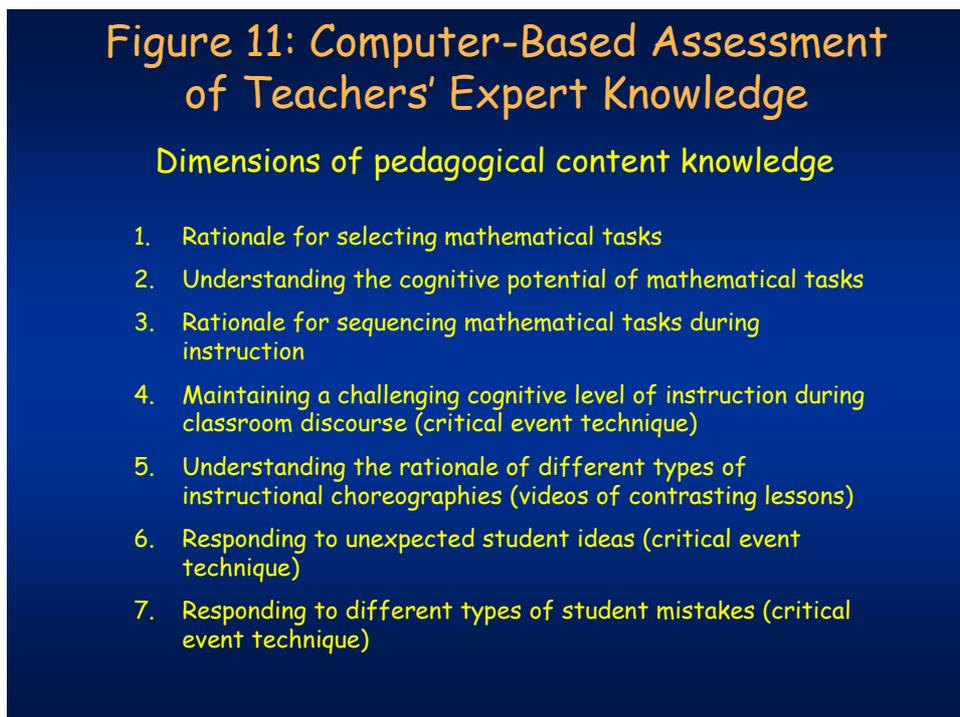
- to show videos presenting critical situations from mathematics lessons or different choreographies of mathematics instruction.

In other words, the instrument allows us to generate situations that require professional responses and necessitate the activation of pedagogical content knowledge.

The instrument aims to address the following aspects of mathematics teachers' pedagogical content knowledge (cf. Figure 11):

**1. Selecting tasks for lessons on a standard grade 9 and grade 10 mathematics topic.**

A series of tasks with varied cognitive potential are presented on the computer screen. They include tasks allowing different levels of student prior knowledge to be taken into account in that they permit several solution-finding processes. Teachers are asked to select appropriate tasks for introducing a new mathematics topic as well as for organising practice periods and to give reasons for their choice.



**2. Handling selected tasks in classroom instruction.**

The aim here is to determine whether the teachers put the cognitively activating potential of tasks to good use.

**3. Sequencing of tasks.**

Teachers are asked to select tasks suitable for introducing a new mathematical topic, to arrange these tasks in appropriate order on the computer screen, and to give reasons for the sequence chosen. In a second step, they are presented with other sequences, each of which represents a different instructional approach. They are asked to select one of these sequences for their lessons and to give reasons for their choice.

**4. Structuring the content of instructional processes at critical points in the lesson.**

Teachers are shown short videos of mathematics lessons. The videos break off at a critical point at which it is determined whether the potential level of cognitive activation is maintained or the instruction is trivialised. The teachers are asked to continue the lesson at this point.

**5. Sensitivity to different instructional choreographies.**

Teachers are presented with four short videos showing examples of different instructional choreographies in the mathematics classroom. The videos, which recreate authentic situations from mathematics lessons in Germany, Japan, Switzerland and the USA, portray the introduction of a new topic at the beginning of the lesson. As soon as the topic has been introduced, the video breaks off. The teachers are asked to continue the lesson and give reasons for their decisions.

**6. Responding to unexpected student ideas.**

Teachers are shown short videos of lesson situations in which students present unexpected mathematical ideas. The teachers are asked to respond to these ideas.

**7. Responding to student errors**

Finally, the teachers are presented with videos of classroom situations in which students make mistakes. These errors represent temporary misunderstandings, typical misconceptions in the topic area, as well as more unusual errors that can be put to good use in cognitively activating instruction. The teachers are asked for their reactions to the errors, which are presented in random order. Again, they are asked to give reasons for their responses.

This computer-based instrument for the assessment of teachers' expert knowledge is complemented by an analysis of randomly selected homework assignments and informal tests. The written material is analysed from the perspective of its implicit pedagogical conception. Finally, a traditional questionnaire is used to assess socio-demographic teacher variables and aspects of the teachers' occupational career, professional commitment and belief systems.

Parallel to this teacher study, we also tap into student expertise. A student questionnaire collects data on those characteristics of the instructional process for which students reports are reliable and have predictive validity. For the purposes of construct validation, these aspects of instruction are also assessed from the teachers' perspective in the context of the traditional teacher questionnaire.

Following initial testing in face-to-face situations in which teachers commented on the computer presentations, the first field trial with 80 mathematics teachers who also participated in the PISA 2003 field trial is now in progress. We do not yet know whether the new assessment format will prove successful as analytical tool. One thing is already certain, however: Our new approach does succeed in involving teachers in an expert conversation – and one that the majority of participants find fascinating. The challenge facing us in the coming months will be to streamline and optimise the instrument, and – in parts at least – to increase its level of standardisation. For the time being, we have shelved our original idea of developing a self-explanatory instrument that could, in the long run, be administered over the World Wide Web. We will be content if we manage to develop a parsimonious instrument in CD-ROM format that can be implemented in Germany in the PISA 2003 main study and in the national longitudinal assessment the following year. The instrument is to be administered by the same test administrators who are responsible for conducting the student assessment. If we succeed in this endeavour, the new instrument may pave the way for the inclusion of a PISA Teacher Study as an international option in the PISA 2006 or 2009 assessment.