Journal of Science Teacher Education (2005) 16: 309–349

Feature Article

Framework for Empirical Research on Science Teaching and Learning[1]

Hans Ernst Fischer, Klaus Klemm, Detlev Leutner, Elke Sumfleth, Rüdiger Tiemann, & Joachim Wirth
University Duisburg-Essen, Essen, Germany

In view of the research on education—and subject-related education in particular—that has been conducted in recent years, it would seem useful to describe the current state and future trends of research on science teaching and learning. In the present article, research findings are described, the deficits of science education are analyzed, and medium- and long-term research goals are specified from the perspective of an interdisciplinary cooperative effort between specialists in the fields of empirical educational research; the psychology of learning and instruction; and biology, chemistry, and physics education.

Introduction

International comparative surveys of student achievement have highlighted weaknesses of German and American science education and—given the normative goals of the respective consortia: the Progress in International Reading Literacy Study (PIRLS, primary school), the Trends in Mathematics and Science Study (TIMSS, grades 8 and 12) organized by the International Association for the Evaluation of Educational Achievement (IEA), and the Program for International Student Assessment (PISA, grade 9) initiated by the Organization for Economic Cooperation and Development (OECD)—revealed an urgent need for further research into how teaching processes can best be structured to harmonize with learning processes (Baumert, Klieme, et al., 2001; Bos et al., 2003; Organization for Economic Cooperation and Development, 2003). Consequently, insights are required into the state of education systems and the organization of individual schools, the conditions and forms of instructional design, the effects of inservice training, and individual learning conditions. Primarily, descriptive research is needed to shed light on these issues and their interrelations. Parallel to this, the insights gained can be implemented in targeted intervention measures and tested empirically. The results of this prescriptive intervention research should then lead to implementation research in all three

---

[1] Revised and supplemented version of Fischer, H. E., Klemm, K., Leutner, D., Sumfleth, E., Tiemann, R., and Wirth, J. (2003). Naturwissenschaftsdidaktische Lehr-Lernforschung: Defizite und Desiderata [Natural science-didactical learning research: Deficits and desiderata]. *Zeitschrift für Didaktik der Naturwissenschaften*, 9, 179–208.

phases of teacher training—more theoretical training at university, practical training in schools, and inservice training.

PISA 2000 was an internationally standardized assessment of reading literacy developed and carried out by 43 countries and administered to 15-year-olds in schools. The second cycle in 2003 focused on mathematics; the third in 2006 will focus on scientific literacy. The PISA tests, which are typically administered to 4,500–10,000 students per country, examine cross-curricular competencies rather than mastery of the school curriculum. Within each domain, items are designed to assess students' command of processes, their understanding of concepts, and their ability to function in various situations. The test items are a mixture of multiple-choice items and questions requiring students to construct their own responses. In addition, students answer a background questionnaire, providing information about themselves and their homes, and school principals answer a questionnaire about their schools. The aim is to produce a basic profile of knowledge and skills among 15-year-old students, contextual indicators relating results to student and school characteristics, and trend indicators showing how results change over time, thus providing a valuable knowledge base for policy analysis and research (see http://www.pisa.oecd.org/pisa/summary.htm). Students in Germany performed well below the Organization for Economic Cooperation and Development (OECD) average in PISA 2000 and 2003; the performance of students in the United States was around the OECD average (Baumert et al., 2001; Organization for Economic Cooperation and Development, 2001, 2003). The TIMSS assessment was the first in a 4-year cycle to evaluate trends in students' mathematics and science achievement. The results, published in 1997, described science performance at the end of the 8th and 12th grades and revealed that 20% of 8th-grade students in the U.S. and Germany do not have the physical knowledge expected of elementary students and that only 25% have an incipient understanding of scientific concepts and processes (Baumert & Lehmann, 1997; Mullis, Martin, & Gonzalez, 2004). One of the remarkable features of both large-scale assessments is their high quality in terms of objectivity, validity, and reliability. Each step of the assessment has been and will be controlled by an international consortium. The national project managements are required to keep a detailed protocol of data collection, and quality control and unannounced checks are an integral part of the study design.

Bearing in mind the literacy concept underlying TIMSS and particularly PISA ("more or less domain-specific knowledge, skills, and strategies that can, in principle, be learned"; Baumert, Stanat, & Demmrich, 2001, p. 22, own translation), the implications of these findings are painfully clear. Both TIMSS and PISA focus on measuring students' ability to apply knowledge and skills in a variety of situations. This move away from the concept of curricular validity (particularly in PISA) and toward the application of scientific explanation and argumentation skills reflects a growing awareness that a lack of skills in specific domains can have repercussions going far beyond the school gates and impacting, for example, on initial vocational training. Germany's better performance in PIRLS (TIMSS for 4th grade) suggests that the real difficulties do not set in until after the 4th grade. By the end of the 8th grade, Germany and the U.S. rank in the broad midfield of the TIMSS international science scale. The equal-sized performance gains at *Gymnasia* and *Hauptschulen* indicate that—from the 7th to the 8th grade, at least—science learning is not cumulative in German schools. Moreover, some 20% of students are only able to solve tasks at the lowest proficiency level (everyday knowledge), meaning that they have learned very little in their 5th–8th-grade science lessons.

The discussion triggered by TIMSS and PISA has led to moves to develop a quality-assurance system defined in terms of performance targets. In this context, knowledge of subject matter is no longer a resource to be employed, but the desired outcome of teaching. With regard to this change in perspective in the control mechanisms of the education system—from describing resources to evaluating the effects of instruction—Lange (1999) and Helmke (2000) have written of a paradigm shift in German educational research, reflected in a shift from input- to output-driven management (see also Klieme, 2003). In the past, the German education authorities' management philosophy was characterized by a belief that the desired effects on cognitive and social learning (Bellenberg, Böttcher, & Klemm, 2001, p. 20) could be achieved by exerting the necessary control on input variables (provision of teaching staff, teacher training, student ability, school facilities, etc.) and process variables (curricula, timetables, teaching and learning materials, rules governing student retention and promotion, etc.). This approach dominated throughout the country, though growing numbers of the 16 federal states (education policy is a matter of the federal states in Germany) sought to incorporate elements of output-driven management in their centralized school-leaving examinations (Klemm, 1998). During the 1990s, however, the effectiveness of the established input- and process-driven approach was increasingly called into question—first, by the development of semiautonomous schools, which was pursued to varying degrees in all federal states and, second, by the results of national and international comparative surveys (van Ackeren & Klemm, 2000). Consequently, school policymakers and administrators began to focus more strongly on output-driven management, and this, in turn, prompted the paradigm shift diagnosed by Helmke (2000) in school research. As the large-scale assessment studies were prepared and the wealth of empirical data analyzed, empirical school research, in general, and the branch focusing on the outcomes of schooling, in particular, became more relevant than ever before.

As a theoretical basis for optimizing science education in line with the TIMSS and PISA objectives, which correspond broadly with the general goals formulated in the curricula of the individual German states, we propose a combination of the quality-assurance-oriented framework model devised by Baumert et al. (2001), the basis models used by Oser and Patry (1994) and Oser and Baeriswyl (2001) to classify instruction, and the concepts of cognitive psychology. A deficit analysis on the basis of this theoretical framework identifies five key areas of concern: (a) the importance attached to science education in society and schools, (b) teacher training, (c) the conception and implementation of (basic) scientific education, (d) horizontal and vertical integration of subject matter, and (e) individual teaching methods specific to the science subjects.

Theoretical Frame of Reference

A Framework Model for the Analysis of Student Performance

In view of the broad range of context factors impacting on school learning, Baumert et al. (2001) and Baumert, Artelt, Carstensen, Sibberns, and Stanat (2002) have taken a framework model for the analysis of student performance as a basis for the PISA reports. This model, which summarizes and systematizes the current state of research, closely following Helmke and Weinert (1997), catalogs the conditions that impact on learning and performance outcomes (Figure 1). The strength of this theoretical framework model is that it does not focus either on out-of-school and institutional factors—as was long standard practice in German school research (see Helmke, 2000)—or on instructional processes, as was the emerging trend in educational science following publication of the TIMSS results. Rather, it covers out-of-school, institutional and instructional conditions impacting on performance and learning outcomes, as well as the interrelations between these factors. The present article adopts this systematization and the underlying theoretical assumptions. It is particularly important to note that the findings of the international studies mentioned above cannot be directly related to the classroom context, instructional processes, or teachers' professional backgrounds.

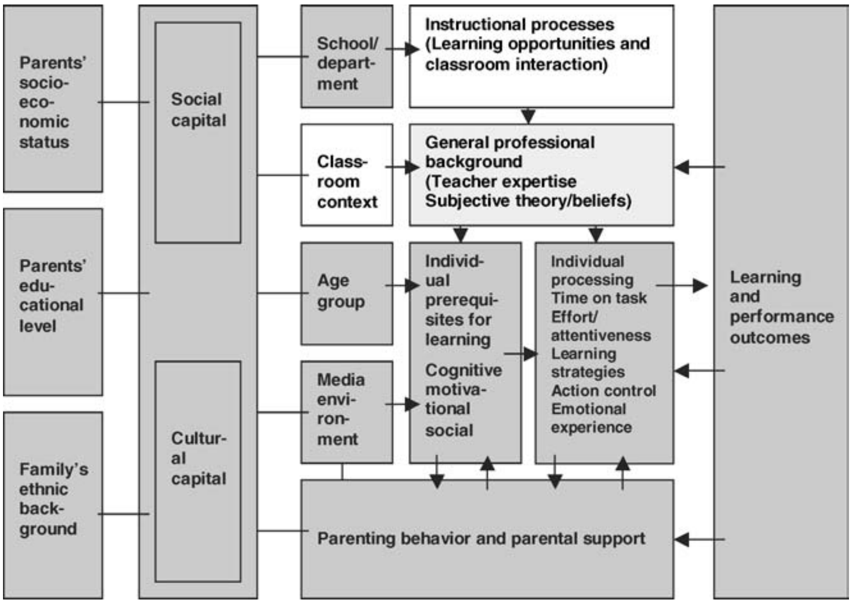

Figure 1. *A framework model for the analysis of student performance (Baumert et al., 2002, p. 16).*

The Basis-Model Theory

The basis-model theory developed by Oser and Patry (1994)—an approach entailing the initially non–subject-specific assessment of instructional strategies—serves as a concrete frame of reference for research on instructional processes (learning opportunities and classroom interaction), as well as on the teachers' general professional background (teacher expertise, subjective theory and beliefs) and on classroom context (see Figure 1). According to this theory, teachers use a limited number of basis models to structure the elements of a lesson, dependent on their teaching goals (Oser & Baeriswyl, 2001). Oser assumed that these basis models can describe teachers' classroom behavior almost exhaustively. According to Baumert and Lehmann (1997), few types of teaching goals are actually pursued in mathematics classrooms. The same patterns of action regulation or, to use the term coined by Oser and Patry (1994), the same *choreographies of teaching* are, in fact, observed again and again. Stigler, Gonzales, Kawanaka, Knoll, and Serrano (1999) confirmed that there is little variation in the instructional choreographies of different mathematics teachers within the three cultural regions of Germany, Japan, and the U.S. The concept of "cultural scripts" used in the TIMSS Video Study (Stigler et al.) to describe the typical surface structure of different lessons seems to have been developed from a rather hermeneutic and descriptive perspective. There has, as yet, been little empirical investigation of the scripts typical for the science classroom (Klieme & Schecker, 2001). Thus far, two research groups have detected a similar trend in physics teaching (Fischer, Reyer, Wirz, Bos, & Höllrich, 2002; Seidel et al., 2002). Following Oser and Baeriswyl (2001), teachers' concepts of instruction can be described as latently effective concepts about teaching and learning. However, it remains unclear whether and how their (few) different basis concepts affect the learning activities of students. According to Reyer (2004), the following basis concepts are relevant for physics teaching: (a) learning by experience, (b) conceptual change, (c) problem solving, (d) development of theory, (e) learning of strategies, (f) training of routines, (g) top-down learning, and (h) learning to negotiate. Only the concepts of learning by experience and development of theory have been observed to be applied by the teachers investigated, however (Fischer et al., 2002). To analyze the effect of basis concepts, it is necessary to assess how they are represented in lessons and whether or not students react in a specific way if they are varied—and, if so, how. If basis concepts are understood as teachers' cognitive schemata that regulate activities in certain classroom teaching situations, their effect on lesson design can be analyzed by categorizing and analyzing video sequences of lessons accordingly (Jacobs, Kawanaka, & Stigler, 1999). If they affect students' behavior and performance, we should also be able to discern them by analyzing lessons. Their effect on students' performance can be analyzed by testing; interest and motivation can be assessed by means of questionnaires. The results of these three different methods of investigation then have to be triangulated.

To investigate empirically this combination of teachers' concepts, lesson structure, and student performance, we need an approach that takes account of teachers' ideas, as well as their activity-regulation cognitions, the representation of these cognitions in lessons, and the related student activities. Determining which level of instruction has the main impact on student learning and performance is a key issue in instructional research. Thus far, neither general nor subject education has been able to provide a great deal of empirically founded evidence. Yet, the call for better science education has inevitably created a demand for criteria of instructional quality and for quality standards that are, at the very least, empirically plausible. If instructional quality is to be defined in terms of learning outcomes, it must be possible to describe these outcomes as a function of teaching and learning processes. Only then can learning outcomes be planned. Several studies have shown that an understanding of instruction on the superficial level of the TIMSS Video Study description is a necessary, but not sufficient, condition for comparing instruction. Oser and Patry (1994) (see also Fischer et al., 2002; Oser, Patry, Elsässe, Sarasin, & Wagner, 1997; Seidel et al., 2002) described the surface structure of a lesson (i.e., the visible structure of teaching) as its sight structure. This can, in principle, take an almost unlimited number of forms. Teachers tend to begin planning their lessons (deciding on student groupings, learning media, tasks, etc.) at this descriptive level. However, descriptions of this surface structure do not sufficiently explain an essential aspect of instructional quality: how instruction impacts on student performance.

The distinction between the surface (or sight) and deep (or underlying) structure can be adopted from Chomsky's transformational grammar (e.g., Chomsky, 1965). The surface structure is regarded as a product of the deep structure (in Chomsky's theory, the latent rule system of the syntax). On the one hand, the deep structure can be viewed as the teacher's latent rules for the planning, design, and conduct of a specific lesson; it "generates" classroom processes in the same way as grammatical syntax generates the wording of a sentence. On the other hand (expanding on the original concept), the deep structure can also be seen in terms of the latent steps in the learning process, that is, the learners' response to the instruction offered. The latent rules of the deep structure prompt the systematic translation of the surface structure into student learning processes. For teachers, the deep structure comprises specific teaching goals, more general intermediate goals, and instructional design; for students, it comprises the successive cognitive steps involved in learning about a particular subject, which converge to form a learning path leading to specific knowledge and skills. The latter are dependent on both the form of the learning path and the content of instruction.

According to Oser and Patry (1994), the deep structure of a lesson can be described—dependent on the teaching objectives—in terms of a limited number of basis models, each consisting of a defined sequence of students' activities. It is not yet clear how the instructional choreographies that evolve from these operation chunks affect student learning activities and outcomes. According to Fischer et al. (2002), however, very stable, more or less successful instructional profiles can be observed to regulate the behavior of different physics teachers in the form of intuitive basis models. Brouër (2001) reported that basis-model oriented instruction in German as the native language seems to have a positive effect on learners' perception and differentiation of learning processes.

For physics, it has been shown that different phases of instruction (e.g., appropriating a new physical concept or drilling a mathematical procedure used to process data) may have a similar surface structure, but different deep structures. It follows that lessons given by different teachers with the goal of appropriating a physical concept may—on the superficial (visible) level—seem to be organized along similar lines, but, in fact, derive from quite different teacher intentions. This explains why lessons with a similar surface structure (summed up as cultural script) can have different effects on student performance and interest, as was found in the TIMSS Video Study (Baumert & Lehmann, 1997; Stigler & Hiebert, 1997). In order to be able to gauge the effectiveness of these basis models or of alternative models, it is therefore necessary to assess not only how the models are reflected in teachers' classroom practice, but also how students are observed to respond to it. Furthermore, if these models are regarded as cognitive schemata for action regulation, their effects on instructional design can be investigated by operationalizing and analyzing teacher and student behaviors in videotaped lesson scenarios (Jacobs et al., 1999), and their effects on student performance and interest can be gauged by administering appropriate tests.

Cognitive Psychological and Constructivist Approach

The cognitive psychological approach, or moderate constructivism, currently prevalent in research on learning and instruction (Fischer, 1993; Labudde, 2000; Leutner, 1997, 1998a) would provide a general frame of reference for investigating individual prerequisites for learning (e.g., cognitive, motivational, and social conditions of the students), individual processing of students (e.g., time on task, effort and attentiveness, learning strategies, action control, and emotional experience), and learning and performance outcomes (see Figure 1).

Whereas behaviorists hold that individual behavior can be steered by external stimuli and reinforcement and take this approach to teaching strategies used for controlled practice, in particular, but also for conveying new material (Baumgartner & Payr, 1994; Leutner, 1992, 1998b, 2002; Mandl & Hron, 1989; Weidenmann, 1993), advocates of the cognitive theory-driven approach regard the learner as an information-processing individual who processes external stimuli actively and independently. Learners are assumed to perceive, interpret, and process these stimuli selectively in accordance with their previous experience and current stage of development. This is reflected in the system of patterns and schemes of perception, comprehension, and processing available to the learner and constituting his or her cognitive structure (e.g., Euler, 1994; Mayer, 2001). Learning is understood to be a generative process during which new information is selected, organized, and integrated (Sumfleth, 1988; Wittrock, 1989). The development of learning environments is thus based on deliberations as to which learning processes could or should evolve from the interaction of teaching material (as an external learning condition) and the cognitive structure (as an internal learning condition). Although cognitive theory-driven approaches attach great significance to individual processing, they unswervingly assume an interaction between external presentations and internal processing. In other words, they assume that learning can be stimulated, supported, and, to a certain extent, steered by instruction and learning aids.

Constructivists are much more skeptical about the possibility of stimulating, supporting, and steering the learning process. As they see it, individuals structure situations self-referentially, construct individual meaning from these situations, and actually contribute to creating them through their perception and interactions. Even empirical knowledge is considered to be a subjective construction of reality to begin with, though it may lead to the social construction of reality through linguistic processes of communication (Glasersfeld, 1995). The implication of all this for science learning is that learning environments cannot be regarded as a means of steering learning processes directly, but, rather, as nonobjective hindrances to self-directed learning (Euler, 1994; Fischer, 1993; Labudde & Pfluger, 1999). In the sense of self-determined, reflective practice, particular importance is attached to the self-regulation of the learning process. The radical constructivist approach, which completely denies a reality outside the individual and thus negates any effect of instruction on the learning process, is disputed, however.

Merrill (1991) has coined the term *second generation instructional design* to describe a pragmatic middle position approaching moderate constructivism. On the one hand, this approach, like constructivism, emphasizes the importance of learning in problem- or action-driven contexts; on the other hand, it assumes—in line with cognitive theory—that the construction of cognitive structures or mental models can be influenced by suitable learning environments (Merrill, 1991; Weidenmann, 1993). From the learners' perspective, the focus is on finding the optimal balance between construction and instruction; from the teachers' perspective, it is on finding a balance between self- and other-determination as regards the adaptive design of instruction and the learning environment (Leutner, 1992, 1998a, 2002; Sumfleth, Wild, Rumann, & Exeler, 2002). Situated learning and authentic learning environments, the approaches favored by Mandl, Gruber, and Renkl (1995) and Roth (1995), can also be regarded as a connecting link between cognitive theory-driven and constructivist approaches. Here, the focus is on the life and learning situation, account being taken of additional motivational and communicative aspects, and the complexity of the learning environment (Fischer & Horstendahl, 1997; Krapp, 1993).

Factors Impacting on Science Education

The function of exercises and tasks in the science classroom, teachers' ability to ensure integration of curricular content, subject-dependent individual learning processes and the function of experiments in the science classroom are of particular interest when it comes to investigating processes of instruction and learning at school in more detail (see Figure 1). In their function as agreed basic descriptions of teaching and learning processes, "instructional scripts" are a particularly promising approach to international comparisons of science instruction. By comparing basic classroom processes on a cross-country basis, we hope to identify potentials for developing subject-specific teaching strategies and teacher training.

Instructional Scripts

The TIMSS Video Study has shown that mathematics lessons in Germany, Japan, and the United States follow culturally specific patterns. Seidel et al. (2002) and Fischer et al. (2002) have analyzed the surface structure of physics lessons and ascertained that students are actively involved in, at most, 15% of classroom activities. Both studies identified two types of instruction: teacher-centered instruction in which teachers demonstrate experiments and teacher-centered instruction in which students are allowed to perform some experiments. The first type of instruction (see also Baumert & Köller, 2000, for upper secondary level) is dominated by classwork, demonstrations of experiments, boardwork and note taking (>80%). In the second type of instruction, about 20% of lesson time at the lower secondary level is set aside for student experiments (rarely for theoretical work). In both cases, lessons are dominated by the teacher, who directs the classroom interaction by a tightly controlling series of questions, and learners participate actively in only a very small proportion of this communicative activity. More attention does appear to be paid to student comments in the teacher-centered phases of instruction involving student experiments. Overall, however, teachers tend not to make student beliefs explicit and, thus, rarely take these beliefs into account (e.g., Sumfleth & Pitton, 1998). This also applies to seatwork, where student–teacher interactions play only a minor role. Instruction involving student experimentation does not focus on the learning process, but on the results of the experiment. Evidently, experimental instruction in chemistry and physics (data are not yet available for biology) is geared to reproducing facts, rather than to applying theoretical knowledge to problem-solving or experimental situations.

Thus, experimental phases of instruction that are not embedded in a broader scientific context in terms of content and structure (vertical integration) and that do not involve planned discourse can, at best, foster manual skills with very specialized apparatus and the ability to follow a series of instructions (Hucke, 2000; Hucke & Fischer, 2001; Lunetta, 1998). In response to the deficits outlined, the effects of science teachers' professional knowledge and instructional behavior in experimental contexts are to be examined within instructional structures that have already been explored (Fischer et al., 2002; Sumfleth et al., 2002). Empirical findings can only be effectively translated into intervention measures in follow-up studies if the framework conditions for science teaching—and particularly teacher and student experiments—are known and if the interactions between these conditions and instructional quality can be described. To this end, it is first necessary to consider the more normative societal preconditions.

Basic Education in Science

A basic education in science is—or is, at least, supposed to be—an integral part of the educational provision of German schools offering a general education. General education (Heymann, 1996; Tenorth, 1994) comprises a society's efforts "to convey to the upcoming generation via its social institutions the knowledge and skills, attitudes and outlooks, a command of which is considered necessary and indispensable" (Tenorth, p. 7, own translation). It is against this background that the norms of basic science education are currently being discussed. The discussion is characterized by the normative approach that has been taken in the U.S. debate (Ledermann, 2001) and that is now spreading to Europe (Gräber & Bolte, 1997; Gräber, Nentwig, Koballa, & Evans, 2002). The general goals of science education have been operationalized by Bybee (1997; see also Bingle & Gaskell, 1994; Fischer, 1998; Glynn, Muth, & Denise, 1994; Jung, 1970; Organization for Economic Cooperation and Development, 2001; Prenzel, Rost, Senkbeil, Häußler, & Klopp, 2001). According to this operationalization, science education should provide students with an understanding of the fundamental phenomena that can serve as a basis for scientific systematization and the development of scientific concepts. The historical roots of the discipline should be elucidated, and the rudiments of the theories of science and knowledge should be presented (in keeping with the learner's age). Scientific reasoning should be both an organizational principle and an instructional goal, with discourse and theory building being viewed as essential components of this approach. Furthermore, presentation and argumentation skills should be fostered so that learners are able to participate in the public debate on current issues and future planning when scientific concepts are involved.

In line with U.S. standards (American Association for the Advancement of Science, 1989; Biological Science Curriculum Study, 1990; National Science Teacher Association, 1990), instruction is to be designed in such a way that students are given the opportunity to understand the specific structures of the natural sciences; come to grips with the subject matter; and apply models, theories, and methods properly. This includes the capacity to make observations and draw the appropriate conclusions and the awareness that all observations are selective and subjective. They should be able to apply scientific methods of measurement and proof; assess the scope and reliability of these methods; work with scientific classification systems; recognize scientific endeavor as a form of social enterprise (research, development, production, application); and become familiar with the causes, processes, and implications of science—while assessing the effects of these on various areas of life. Last, but not least, learners should be made aware that the natural sciences are already, and will increasingly become, essential in many occupational fields. Discussion of these elements, subsumed under the English term *scientific literacy*, has recently intensified in Germany, as reflected in the new science guidelines for comprehensive schools in North Rhine-Westphalia: "Along with the other subjects, science education aims to equip students with skills that facilitate life-long learning, social participation and a feeling of shared responsibility, and individual self-development" (Lehrplan NRW, 1999, p. 6, own translation). Similar demands, with differing points of focus, have repeatedly been made of science education (IPN Curriculum: Frey & Achtenhagen, 1975; Salters: Millar, 1993; STS: Solomon & Aikenhead, 1994; Nuffield: Waring, 1997; Chemistry in Context: Parchmann et al., 2001).

Considering the German and U.S. TIMSS results at the lower secondary level, the normative demand for science education to equip learners with this kind of scientific literacy might seem rather utopian from the empirical point of view (Baumert & Lehmann, 1997; Beaton et al., 1996; Shamos, 1995). Nevertheless, it is important for the debate on educational standards to take place and for curriculum design to be based more firmly on the findings of empirical instructional research. As proposed in the expert report, "The Development of National Educational Standards in Germany" (Bulmahn, Wolff, & Klieme, 2003) and in line with Weinert (2001), competency models are to be developed by opening up a new perspective on curricula. These competency models can be translated into operationalizable teaching goals. Successive competency levels will be distinguished, making it possible to substantiate the learning outcomes expected of students of various ages, describe the step-by-step development of knowledge, and evaluate learning outcomes using appropriate assessment procedures (Weinert, 2001, p. 59). This kind of output control is new to Germany and calls for a new outlook to be taken on school performance. Rather than stipulating minimum or maximum standards, a mean anticipated level of performance (given the variance that can be caused by instruction) is specified as the target level. All this places certain demands on the system: Schools must create conditions that give as many learners as possible the chance to achieve the set goals in the time available, and quality-assurance measures must be implemented to ensure the system achieves the best possible results. From the research perspective, it will be necessary to determine how the normative goals of basic science education can be elaborated in competency models and implemented effectively in the classroom and how this implementation can be controlled and evaluated. Isolated content areas will have to be vertically integrated by means of overarching concepts, and assessment instruments will have to be developed for quality-assurance purposes.

Integration of Content

Competency-level models provide a basis for the targeted development of better integrated curricula, ensuring that the knowledge acquired in the classroom can be slotted into an existing conceptual network. In this context, and against the background of cumulative learning, the main emphasis tends to be on vertical integration, though the level of horizontal integration in Germany also leaves much to be desired. Granted, students only have access to a limited amount of prior knowledge (Weinert, 1996). Nevertheless, overarching scientific concepts can help to structure and elaborate content areas (conservation laws, energy transformation, continuum–discontinuum, the donor–acceptor principle, chemical balance, etc.) and—in the same way as hypertext learning (Oser & Baeriswyl, 2001; Wagner, 1999)—open up various perspectives on a given subject area. Indeed, these overarching concepts can serve as extrinsic advance organizers (Ausubel, 1963) that foster cumulative learning. Seel (2000, p. 262) called them *anchor ideas*. Scientific reasoning, as described above, could function as an analogous advance organizer on the level of the research method.

Weber (2002) suggested that two kinds of *vertical* integration be considered, both of them consecutive in terms of chronology and content. *Internal* vertical integration refers to the structure of the subject matter covered in distinct content areas (mechanics, heat, etc.); *external* vertical integration refers to the points of interface between these domains. Individual content areas tend to be presented in isolation in school, particularly in physics classrooms. Yet part of the difficulty of chemistry derives precisely from the fact that the different domains are so closely interlinked (e.g., by an understanding of chemical reactions and symbols or of models of atoms and molecules) that learners who have not yet grasped these relationships have no chance of engaging in continued meaningful learning. This strong vertical integration is not apparent in either the 45-min rhythm of the timetable or the chapter structure of textbooks. As yet, there has been only one empirically founded attempt in Germany to develop a didactic concept fostering cumulative learning—and even that was in an isolated area of physics instruction, namely optics (Weber, 2002). Considerable further research is required here.

The effects of *horizontal* integration of different subjects (that may be taught in parallel) are thus largely unknown, although this is an accepted approach that has been implemented, in part, in science education at the lower secondary level in both the U.S. and Germany. One might reasonably assume that it is easier to design and implement domain-transcending, integrative instruction if biology, chemistry, and physics are covered in a general science course than if they are taught separately—provided, of course, that teachers have the necessary training. However, the complexity of problems is considerably amplified in integrated or combined science courses as compared with separate courses for each subject. In chemistry and physics, subjects already characterized as particularly difficult, increasing the level of complexity will certainly not facilitate students' understanding of the respective models and concepts (Schecker & Winter, 1999). General science courses allow a more strongly example-based approach to be taken; separate courses for each subject facilitate a more systematic approach. Accordingly, care must be taken to ensure that the content structure of general science courses is systematic and allows for the necessary differentiation between the various perspectives. In separate science courses, in contrast, it is important to emphasize horizontal integration by giving examples that illustrate the common ground between the subjects. Scientific literacy requires horizontal *and* vertical integration, enabling students to construct cognitive structures onto which new knowledge can be docked (Sumfleth, 1988). The formal and organizational integration of the individual subjects, in a general science course, for example, could also be a first step toward achieving the status of a core subject (Prenzel et al., 2001). Variants specific to the school type should be considered and explored in this context. Horizontal and vertical interconnections can be emphasized using concept maps, for example, as a form of structural aid or learning strategy (Hucke & Fischer, 2001; Leopold & Leutner, 2002; Leutner & Leopold, 2003a, 2003b; Schreiber, 1998; Sumfleth, 1985, 1988; Sumfleth, Bergmann, & Dannat, 1990; Sumfleth & Dannat, 1988; Sumfleth & Stachelscheid, 1986; Sumfleth, Stachelscheid, & Gramm, 1989) or by setting exercises that encourage learners to take new approaches (Fischer & Draxler, 2002).

Exercises in the Science Classroom

Besides experiments, a considerable proportion of science instruction is devoted to exercises. Häußler and Lind (1998) defined exercises as "...well-defined problems with (at least) one solution that can be processed in a relatively short time. Most physics exercises are formulated in words and require an answer to be given in writing" (p. 3, own translation). Exercises are employed in science lessons with various intentions. They may serve to appropriate new knowledge, perform routine procedures or transfer these routines to new fields of application, and their potential to structure learning allows the orchestration (Duit, Fischer, & Müller, 2002) of different phases of instruction. Moreover, they allow teachers to communicate scientific methods, such as generating hypotheses or weighing up evidence, and to organize their instruction according to these principles of scientific reasoning (Fischer, 1999; Labudde, 1993; Mueller & Dweck, 1998; Stebler, Reusser, & Ramseier, 1998).

The suitability of the exercises currently used in science classrooms is as yet unverified. The TIMSS Video Study of mathematics instruction (Klieme & Bos, 2000) and individual studies of physics education (Fischer & Draxler, 2002) have indicated that the exercises employed do not promote such skills as the meaningful application of knowledge; rather, they train students to execute algorithms. This may be a convenient way of drilling solution methods, but it does not foster students' understanding of the underlying concepts. There is a fundamental difference between biology, chemistry, and physics classrooms, however. In physics lessons, seatwork is dominated by exercises (the quality of which is as yet largely uncharted territory as regards cognitive demands and learning effectiveness), but chemistry lessons feature very few exercises. Barely any of them require problem solving in complex situations (Gabel & Bunce, 1994). Zoller (1990) has proposed a computer-based environment for solving complex, everyday problems that involve chemistry-related issues, but this has not led to the development of chemistry-specific exercises.

Thus, it seems that exercises are not employed to their full potential in German science lessons (Baumert et al., 2001). Possible reasons for this were identified by the TIMSS Video Study (Stigler & Hiebert, 1997), in which a total of 231 eighth-grade mathematics lessons in Japan, the United States, and Germany were video recorded. Systematic cross-cultural comparison of the surface structures of these lessons revealed that Japanese mathematics instruction is particularly geared toward promoting problem-solving skills. The exercises set are often open ended, permit different forms of practice and application, and admit various solution methods. They thus differ considerably from those administered in German or U.S. classrooms, which focus on the result, rather than on the solution process or on the discussion of different approaches. Accordingly, a concept for the use of exercises in science education, including the experimental phases of instruction, needs to be developed and tested. On the basis of existing criteria (Fischer & Draxler, 2002; Klieme & Baumert, 2001; Klieme, Funke, Leutner, Reimann, & Wirth, 2001; Leutner, 2002b; Organization for Economic Cooperation and Development, 2001; Wirth, 2003), strategies can be developed for constructing exercises that defer to the learning path and relate to given competencies. These kinds of exercises can be used in various instructional situations (developing new concepts by theoretical and experimental means, drilling routines, diagnosing and giving feedback on student proficiency, and controlling performance and quality).

Several systems have recently been proposed to describe the difficulty level of exercises and the competencies needed to solve them. The workability of the various approaches cannot yet be gauged, however. Klieme (2000) has identified a set of attributes that characterize the TIMSS items in terms of the cognitive demands they make of a "typical" test taker. The instrument covers eight general characteristics of mathematics and physics items: (a) knowledge of definitions and mathematical/physical theorems, (b) qualitative understanding of mathematical and physical concepts, (c) arithmetic, (d) operations using mathematical terms and calculus, (e) interpreting diagrams, (f) text comprehension, (g) spatial visualization, and (h) problem-solving processes. As yet, no analogous classification systems have been proposed for chemistry or biology items.

Based on a detailed analysis of the items administered in the PISA study, Prenzel, Häußler, Rost, and Senkbeil (2002) have identified the characteristics that predict the difficulty of an item. As criteria of analysis, they differentiated among the formal characteristics of the item (e.g., length of stimulus, task contains numbers, or graphical output), the cognitive demands of the item (e.g., process information from the text, establish logical relations, or calculate something) and the characteristics of the knowledge base needed to solve the item (e.g., relevant information found in text, factual knowledge, or comparative relations). The relationships between these task characteristics (predictors) and the item difficulty (criterion variable) were calculated using regression analysis. The authors found that items tended to be more difficult if students were required to calculate something, use terminological knowledge and give extended responses, or construct a spatial model. The provision of graphs and visual information, in contrast, tended to make items easier. The length of the stimulus barely had any effect on item difficulty. Thus, to a certain extent, this analysis of item difficulty validated the principles of task construction underlying PISA (Prenzel et al., 2001). The PISA science items emphasized the following cognitive aspects:

1. drawing the correct information from a graphical representation,
2. retrieving and applying factual knowledge and skills from memory,
3. drawing the correct conclusions from the information provided,
4. applying mental models, and
5. verbalizing scientific content and concepts.

A competency-level model that adopts this approach to task analysis has been proposed in the context of the program on increasing the efficiency of mathematics and science instruction initiated by the joint central and state government commission for educational planning and research promotion in Germany (Bund-Länder-Kommission [BLK], 1997) and quality development and quality assurance of instruction at the lower secondary level (Klieme & Baumert, 2001). This model takes a more differentiated view of problem-solving processes, and defines overcoming misconceptions as the highest competency level. Fischer and Draxler (2001) have reviewed the various requirements of physics exercises articulated in the literature and drawn up a catalog of criteria intended to provide a theory-driven classification of physics exercises, thus facilitating their targeted deployment in the classroom. Essentially, they have distinguished the content area, solution methods, response format, competency levels, and phase of instruction:

- The content area is defined in terms of the topics covered and how realistic this subject matter is for students.
- Solution methods are defined as experimental if the task focuses on conducting and analyzing an experiment and actively engaging with the experimental method. They are categorized as semiquantitative if a graphical representation is to be interpreted or a new one drawn on the basis of data provided. Finally, an approach is considered arithmetical if the exercise is to be solved using a physical law and mathematical skills.

Klieme (2001) estimated the response format to explain some 30% of the difficulty of an item. Fischer and Draxler (2002) distinguished between multiple-choice items, short-answer items and open-ended items. The open-ended response format is of particular interest because the wording of the experimental tasks used in science education is often either very specific or so vague that students have great difficulty in solving them. In the former case, the level of cognitive activation is too low; in the latter, students feel out of their depth owing to the openness of the situation and the lack of structural guidance.

- Following Dörner (1996) and Fischer (1989), Horstendahl, Fischer, and Rolf (1999) differentiated according to the degree of support students are given with experimental tasks and distinguished between imitative experimentation (simply executing a series of instructions), organizational experimentation (setting up and performing an experiment independently), and conceptual experimentation (students themselves are responsible for deciding which variables are to be measured and for planning the experiment accordingly).
- The competency levels identify the level of proficiency needed to solve an item. Fischer and Draxler (2002) distinguished between applying everyday scientific knowledge; explaining scientific phenomena in simple terms; applying laws and factual knowledge; applying concepts, procedures, and models; engaging in scientific argumentation and problem solving; and overcoming misconceptions.
- Finally, the phase of instruction in which an exercise is to be used is considered: appropriation, practice, or evaluation.

As previously stated, this catalog of criteria can be used to characterize existing physics exercises. Corresponding systems for biology and chemistry exercises have yet to be developed.

The PISA results revealed strong correlations between reading literacy and scientific literacy. For the most part, however, the level of reading literacy in the German-speaking countries is less than satisfactory (Baumert et al., 2001), meaning that many students lack the key prerequisite to succeed in science. It is thus important to consider aspects of reading literacy when constructing exercises for use in the science classroom (Fischer & Draxler, 2002; Tiemann, Fischer, Labusch, & Draxler, in press), especially when these exercises are presented in written form.

According to Fischer (2001, p. 42), well-constructed physics exercises give learners the opportunity to identify with a problem, consider socially relevant issues, test and develop their ideas, apply physical models in a controlled environment, and discuss physical concepts. Thus, properly formulated and deployed exercises make it possible to meet the standards of scientific literacy in the classroom (Bybee, 1997; deBoer & Bybee, 1995; Fischer, 1998). Further research is needed into the specifics of this implementation, or proper construction, and into whether and to what extent Germany's disappointing performance in international comparative surveys of student achievement might be attributable to the widespread use of substandard exercises in German science classrooms.

Experiments in the Science Classroom

Although experiments (in the form of teacher demonstrations and student experiments) are an undisputed and integral part of science instruction (Rosen, 1954), there is no more evidence for their effectiveness in advancing the learning process than there is for exercises. Science education in the United Kingdom and the United States has been oriented around student experimentation since the early 1900s, and even more so since the 1960s. Yet in the late 1970s and early 1980s, the function of experimentation as an element of science instruction was increasingly called into question. Woolnough (1983) even went so far as to call for all traditional expectations of experimentation—particularly the notion that learners conducting physical experiments experience conceptual change—to be abandoned. Indeed, many studies have indicated that labwork does not fully achieve its postulated goals (Bates, 1978; Hofstein & Lunetta, 1982; Toothacker, 1983). The relationship between successful learning and the design of the learning environment has been made clear, however (Woolnough, 1983, p. 24). Tamir and Lunetta (1981) criticized the fact that learners are given very little opportunity to discuss or test their own hypotheses, but are expected to follow instructions as if working from a cookbook (Gallagher & Tobin, 1987; Guillon, 1995; Hucke & Fischer, 2001, 2002). Attention is rarely focused on the scientific content of the experiments or on the differences between scientific models and student misconceptions (Champagne, Gunstone, & Klopfer, 1985; Eylon & Linn, 1988). "To many students, a 'lab' means manipulating equipment and not manipulating ideas" (Lunetta, 1998, p. 250). This quote reflects the findings of many recent research projects dealing with labwork in science education (e.g., Millar & Driver, 1987; Rosen, 1954; Tamir & Lunetta, 1981; van den Berg & Giddings, 1992). Stebler et al. (1998) suggested a possible remedy to this situation: They attributed the good TIMSS performance of 7th graders in the German-speaking part of Switzerland to the fact that their curricula not only require that students be able to “apply factual knowledge and problem-solving skills flexibly,” (p. 48) but also to ensure that this requirement can be met by exposing students to “experiments as productive learning tasks” (p. 48)

In Germany, few empirical studies have as yet examined the experimental approach to learning, lent empirical support to efforts to modify this experimental approach, or allowed conclusions to be drawn about the effectiveness of the labwork conducted in schools. As a rule, the literature is restricted to general suggestions for improving instruction or descriptions of new experiments in specific topic areas. Very recently, researchers have begun to investigate the effectiveness of physics labwork at universities (e.g., Theyßen, 2000). Haller (1999) explored the role of learners’ action goals and the effects that changing experimental instructions while controlling for the complexity of the content has on the learning process. Sander (1999) analyzed the learning processes involved in applying a model-building system in open-ended lab with a strong lecture component; Hucke and Fischer (2001) investigated the comparative levels of knowledge acquired in individual, traditional, and computer-based (model-building) labs. In sum, the findings of these studies indicate that experimental instruction only results in the expected competencies if it is planned specifically with these competencies in mind and specifies an appropriate level of openness as an instructional goal, thus allowing students to experience autonomy in experimental situations.

Experimental instruction in all school types can be differentiated into five phases according to the form of scientific reasoning involved: the planning and design phase (formulating hypotheses), the execution phase (experimentation), the analysis and interpretation phase (discussing results), the application phase (working on a new problem, new hypotheses), and the presentation phase (Bund-Länder-Kommission-Expertise, 1997; Fischer & Breuer, 1997). Communication of one’s ideas is thus an important part of the learning processes to be aspired to in the scientific disciplines (Sumfleth & Pitton, 1998; Sumfleth, Ploschke, & Geisler, 1999). It seems reasonable to assume that an instructional approach implementing each of these phases in the correct order stands a good chance of success. Indeed, according to Stebler et al. (1998) and the TIMSS Video Study (Baumert & Lehmann, 1997; Stigler & Hiebert, 1997), it seems that students’ mediocre performance and lack of interest in science can be attributed to substandard instructional scripts and scarce possibilities for interaction in the content area. The inquiry approach, discussed primarily in the context of environmental education in the United States and the United Kingdom, is based on a description of the same deficits of experimental instruction, but only a very pragmatic action-research approach to filling these gaps can currently be discerned (Harland, 2002; Tamir & Lunetta 1981). Therefore, future empirical instructional research should focus additionally on implementing—and exploring the function of—interactive structures in experimental situations. Other areas belonging to this domain of research include the investigation of corresponding instructional structures and exercise-based learning sequences, the exploration of individual learning and problem-solving processes in experimental situations, and the inspection of individual learning strategies used to tackle scientific problems.

Learning Processes

TIMSS and PISA have shown that students in Germany are relatively proficient when it comes to performing routine science exercises, where scientific knowledge is applied to compute an answer (e.g., a physical quantity) by selecting, combining, and performing known operations. This kind of routine procedure does not necessarily result in permanent conceptual growth. According to Hussy (1984), it is thus important to distinguish the process of performing routine procedures from the process of problem solving. The act of reaching a situational goal state is considered to be an *exercise* when the necessary declarative and procedural knowledge is available and can be applied as a matter of routine. In this case, the process of knowledge application does not involve learning processes serving to seek, identify, and absorb new information or to change cognitive structures permanently. Conversely, a task is considered to be a *problem* when the learner does not have recourse to (all of) the necessary declarative and procedural (prior) knowledge, or has to restructure this knowledge before it can be applied. In this case, learners are required not only to modify the given situation, but to discover or appropriate the information needed to solve the problem (e.g., Preußler, 1997). The process of problem solving—as opposed to performing routine exercises—is thus characterized by the additional learning skills it involves: identifying information and structuring one's approach. However, problem solving cannot necessarily be equated with a learning process, inasmuch as the newly discovered information only has to remain available until the problem has been solved and the specified goal state achieved. It does not explicitly require conceptual change to occur with new information being integrated into the existing knowledge structure (the defining characteristic of a learning process). As such, although learning is not automatically excluded from the process, it is certainly not an integral component of problem solving.

Knowledge is not only applied to achieve goal states by routine means, however. Scientific declarative and procedural knowledge can also be applied with the aim of discovering, identifying, and appropriating new information about a particular topic (i.e., of effecting a quantitative or qualitative change in the knowledge base). An instructional approach (successfully) targeting this objective in the science classroom entails generating hypotheses, testing these systematically in experiments, and integrating the new information thus obtained into students' knowledge structures.

According to Klahr and Dunbar (1988), the hypothesis-generating and -testing aspect of learning can be described as Scientific Discovery as Dual Search (SDDS; Dunbar & Klahr, 1989; Klahr, Dunbar, & Fay 1993; van Jooling & de Jong, 1997). Based on the dual-space model developed by Simon and Lea (1974), they assume scientific knowledge to consist of a hypothesis space and an experiment space. Learners' (more or less secure) knowledge about the relations between variables and the effects of any changes in these variables is represented in the hypothesis space. It is here that hypotheses are formulated, modified, evaluated as valid or invalid, and stored. The experiment space, in contrast, contains all the tests that can possibly be performed to confirm or reject a hypothesis in a given situation. In this context, applying scientific knowledge means systematically selecting, planning, and conducting an experiment or a series of experiments, the results of which will shed light on the validity of the hypotheses. This model makes it possible to describe successful strategies for discovering or producing new information about a given (scientific) content area. However, it does not allow conclusions to be drawn about how this information should best be processed and integrated into one's own knowledge structure so that it can be reliably and easily retrieved and applied in future situations.

Research on self-regulated, strategic learning (e.g., Artelt, 2000; Baumert, 1993; Baumert & Köller, 1996; Leutner & Leopold, 2003a; Lompscher, 1994; Schiefele & Pekrun, 1996; Schreiber, 1998) describes the process of integrating newly discovered or generated information, as well as the conditions, characteristics, and measurement of this process. However, this research approach is often limited to cognitive and metacognitive learning strategies and resource-management strategies, the application of which is supposed to guarantee that once identified, information can be reretrieved and reapplied at later points in time. Models of self-regulated learning do not tend to include strategies for identifying or generating the information in the first place.

At any given time in a learning process, the learner must decide whether to pursue the goal of identifying new information or the goal of integrating known information to ensure it can be applied in the future (Wirth, 2003). The regulation of the learning process must constantly adapt to the changes in the learner's knowledge base and strike an optimal balance between identifying and integrating new information at all times. According to Schreiber (1998; see also Leutner & Leopold, 2003a), higher level learning strategies (i.e., metacognitive control strategies) are thus needed to regulate the application of the lower level strategies of identification and integration (Klauer, 1985). Both the SDDS model (Klahr & Dunbar, 1988) and research on self-regulated learning consider just one aspect of (discovery) learning and—taken in isolation—are, therefore, not well suited to describe the entire process of scientific learning and its regulation. Rather, these two research approaches should be fused to provide a comprehensive insight into the processes of independent identification and integration inherent in scientific learning and action.

The Importance Attached to Science Education

In educational science, the debate on what makes a good school (Tillmann, 1989) has tended not to consider the value that society places on science or the resultant anchoring of the discipline within the school system. In schools, the focus was, and continues to be, on activities targeted at improving the school climate, school life, or school profile, but rarely at enhancing instructional quality and, in due course, the learning and performance outcomes achieved. Up to now, school quality development has rarely centered on instructional development (Helmke, 2000). All this is currently caught up in a process of change, however, not least in response to the sobering findings of large-scale assessments. Yet the risk of the paradigm shift described above is that we may lose sight of the significance of school quality for instructional quality, particularly because there has, as yet, been little empirical research into the nature of this relationship (Ditton, 2000). As long as this remains the case, any intervention strategies developed will be on shaky ground. Thus, it will be necessary to consider the anchoring of science education in the school system as a whole, as well as in individual schools, if promising intervention programs are to be devised to improve learning outcomes in science. Each of the following five domains impacts on the outcomes of science education and should be taken into account:

- the embedding of science in the social culture,
- the amount of time dedicated to science education,
- the sequencing of science education,
- the anchoring of science education in individual schools (Fend, 2001), and
- the students' previous experience of science.

In terms of these five domains, the influence of social and cultural capital on individual schools, on the classroom context, on the cohort of students investigated, and on their parents' behavior and support (see Figure 1) is described in more detail below.

The Embedding of Science in the Societal Culture

At the beginning of the nineteenth century, science was shunted to the sidelines of the educational canon (Blankertz, 1985; Fuhrmann, 1999, especially Chap. 15 on mathematics and science). Some authors have attested to the lasting effects of this development, which was triggered by New Humanism: "Although scientific knowledge does not have to be concealed, it does not rank as education" (Schwanitz, 1999, p. 482, own translation). The members of the joint central and state government commission for educational planning and research promotion in Germany (Bund-Länder-Kommission-Expertise, 1997) clearly had such observations in mind when stating that

> *It is possible to admit a lack of aptitude in these domains without losing face—in other words, this does not seem to hamper individual development in any way. . . .Evidently, mathematics and science education has not yet managed to make students aware of the meaning of these subjects. This cannot be achieved in the classroom alone. (p. 69, own translation)*

The Amount of Time Dedicated to Science Education

The curricula are the main instruments used by societies to give school learning "thematic and temporal structure" (Diederich & Tenorth, 1997, p. 81). According to the German state-by-state comparison of PISA results, "The amount of lesson time specified in the curriculum seems to be an indicator for the institutionally anchored significance and value attached to education in a state" (Baumert & Artelt, 2002, p. 233, own translation).

At 10%, the relative proportion of science instruction received by 12- to 14-year olds in Germany is close to the PISA average (11%). However, the absolute volume is just 271 hours, compared with a PISA average of 306 hours. The discrepancy is even more glaring when Germany is compared with other OECD countries: France 336, England 338, and Austria 443 hours (authors' calculations based on Organization for Economic Cooperation and Development, 2001, p. 249). In Germany, the largest cross-state difference in the total number of lessons scheduled from grades 1–9 are 8,076 lessons in Berlin versus 9,240 in Bavaria, which amounts to more than 1 school year. Cross-state differences in the number of lessons per subject can also be expected (Baumert et al., 2002, p. 48).

The Sequencing of Science Education

The effects of the way a subject is anchored in the temporal structure of the curriculum (i.e., the sequencing of a course of education) are of a similar magnitude as the amount of lesson time devoted to a subject. In the gymnasia of North Rhine-Westphalia, for example, biology is taught in grades 5 to the first half of grade 7 and grades 8, 9, and 10. Chemistry is taught in grades 7, 9, and 10; and physics is taught in grades 6 and 8–10. According to the authors of the PISA report,

> *The considerable cross-state differences in the chemistry literacy of Gymnasium students can arguably be attributed to the fact that the subject does not feature on the curricula until late in the school career. "As a rule, chemistry is not taught as a separate subject until the 9th grade(. . .)In Bavarian Gymnasia that do not have a special scientific profile. It is not introduced until the 11th grade." (Baumert et al., 2002, p. 152, own translation)*

The BLK expert report identifies this lack of sequencing in science instruction as a serious obstacle to horizontal and vertical integration (Bund-Länder-Kommission-Expertise, 1997, p. 45). Neither type of integration can be achieved when subjects are not taught continuously or in parallel, but sporadically and in scraps. For physics education in particular, horizontal integration must be expanded to include aspects of mathematics instruction. It is important that these relationships are not overlooked by approaches endeavoring to improve performance in the individual science subjects.

The Anchoring of Science Education in Individual Schools

The internal cultural settings of an individual school (Fend, 2001) comprise the school itself and the traditions that have evolved there. They include, for example, the amount of instruction given by teachers not trained in the subject. To illustrate this point, an analysis of data from the 2000–2001 school year in North Rhine-Westphalia revealed considerable differences across the different school types: In *Gymnasia*, the overall proportion of instruction given by teachers not trained in the subject is 4.8%; the figure for physics instruction is the same. In *Realschulen*, the proportions are 16.3% and 23.1%, respectively; in *Hauptschulen*, 55.6% and 52.8%, respectively. A comparable pattern emerges for the other science subjects (authors' own calculations based on MSWF–NRW, 2001; see also the overview in the German state-by-state PISA analysis, Baumert, Artelt, Klieme, et al., 2002, p. 204). There does not yet seem to have been any empirical investigation into whether head teachers discriminate in favor of—or against—particular subjects when deciding which lessons are to be taught by teachers without the appropriate training. Incidentally, the same applies to the cancellation of lessons owing to staff shortages, illness, and so forth.

Individual school profiles are another very relevant aspect here. The only available representative data on profile building in the German school system were gathered in the context of PISA and show that scientific profiles are less pronounced than school profiles emphasizing vocational and practical learning, new technologies, and music or art (Baumert et al., 2002, p. 441).

Finally, research on effective schools has shown that professional cooperation between teachers is of great relevance to instructional effectiveness. Levels of staff cooperation are often poor, however, and the suboptimal sequencing of science education in German schools has served to exacerbate this tendency in science departments. It is worth noting that, although the significance of an individual school's culture is well accepted in the literature, empirical findings on the domain-specific "dignity" (Bund-Länder-Kommission-Expertise, 1997, p. 99) as regards science education are still outstanding.

Students' Previous Experience of Science

Science instruction at the lower secondary level builds on the knowledge that students are assumed to have acquired at elementary school. Yet the results of a study conducted in the 1990s (Möller, Tenberge, & Ziemann, 1996) have indicated that science has tended to be neglected at the elementary level. Nevertheless, the BLK experts' impression that insufficient lesson time is devoted to the natural phenomena of inanimate nature in elementary schooling (Bund-Länder-Kommission-Expertise, 1997, p. 5) does not seem to be corroborated by a casual glance at the results of PIRLS (Bos et al., 2003). In Germany, this international reading survey was extended to include an assessment of mathematics and science literacy at the end of elementary schooling. The findings are encouraging. On average, the performance of elementary students in Germany is on a par with that of the international leading group in the corresponding international study. However, it is not clear how these results should be interpreted with respect to the individual subjects. School instruction only seems to play a minor role. Indeed, according to expert ratings, only about 37% of the PIRLS items correspond with the German curricula (Bos et al., 2003, p. 161). Moreover, in a study conducted by the authors of the present article, 10 experts rated a total of 58% of the PIRLS items to tap everyday knowledge: 2% of the chemistry items (one item), 32% of the physics items, 4% of the biology items, and 2% of the geography items. We can thus draw the tentative conclusions that the PIRLS assessment is not an accurate reflection of elementary-level science instruction and that elementary students in Germany only acquired a minor part of their knowledge in the content area under assessment at school, most of it having been picked up from the family, peer group, or television. It has not yet been possible to determine the part played by each of these domains. Nevertheless, it can be assumed that the PIRLS items drawing on students' everyday knowledge are located in a competency domain belonging to their out-of-school environment. This is no longer the case at the lower secondary level. As such, the reform of school science education should not focus primarily on the shortcomings of lower secondary schooling, as might have been assumed in the light of PIRLS results. Rather, it should equally, as suggested in the BLK report, "address the amount of scientific content covered and the quality of this coverage at elementary school" (Bund-Länder-Kommission-Expertise, 1997, p. 55).

Teacher Training

In all German states, teacher training emphasizes academic content rather than teaching methods, particularly where aspiring gymnasium teachers are concerned, and future science teachers are given training geared more strongly to the profession of natural scientist than to the teaching profession. The fact that subject didactics previously lacked a well-defined academic profile has contributed to this focus on the content of the subject to be taught. A form of academic training geared to the profession of science teacher has yet to be developed to strengthen the connection between the teachers' general professional background and instructional processes on the one hand and individual and group processes and student learning outcomes on the other (as shown in Figure 1). It seems, for example, that teachers at all levels lack diagnostic and methods skills, though to differing extents (Baumert et al., 2001). Teaching concepts and instructional goals tend to be dominated by the substance of the subject, and to overlook the fact that variations in lesson content only go a small way in explaining the differences in student performance. As such, the normative understanding of "good teaching" is coming under increasing criticism in the debate on the teaching profession, the professionalization of teachers, and the rationale behind certain teacher-training programs. Conversely, the development of a professional identity is expected to occur as part of an iterative learning process in which the individual is exposed to situations involving professional action, justification and decision-making, drawing on the relevant cognitive knowledge, normative orientations, and pragmatic teaching skills (Bauer, 1997; Bauer, Kopka, & Brindt, 1999; Dewe, 2000). Radtke (1996) described the opportunities and constraints of academically oriented self-reflection aiming to professionalize teaching and developed criteria for the overhaul of teacher training. The implications of this for teacher training are that teachers must be given sufficient exposure to situations requiring pedagogical action and decision making, thus giving them the opportunity to integrate their teaching skills with general didactic, subject didactic and subject-specific skills in a reflective manner and to develop these skills systematically. The derivation of professional behavior is essentially normative, and the ability to mediate between "professional knowledge" and "disciplinary knowledge" is seen as a necessary precondition for successful teaching (Dewe, 2000). According to some authors (e.g., Kennedy, 1994; Moallem & Earle, 1998), successful teaching is dependent on such standards as the ability to assess systems, to apply selection criteria relating to content areas and learning theory, or to base instruction on epistemological considerations.

Many proposals for inservice training programs have been influenced little by the discussion on the quality of instruction. Reports on their effectiveness are contradictory. Pallasch (1997) described instructional supervision as a form of ongoing vocational guidance for those working in education. The Kiel supervision model was described in Pallasch, but its outcomes were not evaluated. Klippert (2000) proposed ways of introducing new teaching methods to the classroom and opened up new forms of interaction with the school environment (consensus building among staff, reducing the workload through teamwork, subject methodology, sensitizing parents, etc.). In their empirical study on the effectiveness of verbal and video-based feedback in teacher training programs, Borchert, Dahbashi, and Knopf-Jerchow (1996) found that procedures aligned to real classroom situations yielded better results. Teacher training should thus be dovetailed with classroom practice and based on sound empirical theory. The Constance Training Model (Konstanzer Trainingsmodell, or KTM; see Tennstädt, 1992), intended as a self-help program for teachers in all school types and grades, consists of a set of documented techniques and practice materials; and it is one of the few teacher training programs that have been given a positive evaluation. Although the KTM was developed to help teachers deal with conflict situations (aggressive student behavior and discipline problems, in particular), its theoretical foundation allows it to be transferred to other areas. The KTM approach displays some parallels with the approach taken by Fischler and Zedler (1999). Because the latter project (part of the DFG priority program on the Quality of Education: BIQUA) is still in the piloting stage, however, it is not yet possible to gauge its effectiveness.

Generally speaking, inservice teacher training programs can be classified into three groups. The first focuses primarily on broadening and differentiating the available repertoire of teaching skills and methods. Training modules propose specific, tested ways of introducing new methods of instruction by taking into account the school environment, the importance of consensus building among staff, sensitizing parents, and so forth (e.g., Borchert, Dahbashi, & Knopf-Jerchow, 1996; Klippert, 2000; Korte, 1998). A second group of concepts is more concerned with supervisory mentoring, be it external evaluation, internal supervision of everyday classroom practice by one's colleagues, or both (e.g., Pallasch, 1997; Pallasch, Mutzek, & Reimers, 1992). The third group pays particular attention to the institutional and system-related aspects of organisational and staff development in the school-development process (e.g., Horster & Rolff, 2001; Kempfert & Rolff, 1999).

Clearly then, in addition to instructional development in the narrow sense, it is important to consider organisational and staff development, or the development of the professional self (Bauer, 1997; Klippert, 2000; Rolff, Buhren, Lindau-Bank, &

Müller, 1998, p. 13). In other words, apart from individual influences, it is vital to consider factors located at the level between the school as an organization and the level of individual action. These factors impact on cooperation, management and leadership, and staff motivation. The practicalities and effects of consulting external advisors, moderators, or trainers are also discussed in this context.

Although these concepts are certainly to be seen in a positive light, it should be noted that they address specific classroom behavior at the level of the surface structure. Furthermore, they have not yet undergone objective evaluation with respect to their effects on student achievement gains or teachers' concepts of good instruction. In the domains of instructional development and professional knowledge, it is generally the case that professional knowledge is placed in opposition to disciplinary knowledge—the two domains of knowledge are considered separately and, yet, expected to converge in teacher behavior. This corresponds to the gap currently found between the subject of instruction and what is known as *Grundwissenschaften* or basic science (the study of social science, including pedagogy) in Germany and beyond. First attempts to integrate the two domains have been made through the cooperative efforts of individual didactics experts, content specialists, educational scientists, and psychologists of learning and instruction. From the integrative perspective, it is not a question of merging the two domains, but of interpreting teachers' specific, disciplinary knowledge of a subject to be a *part* of their professional knowledge. In concert, these two domains constitute the core curriculum.

There has, as yet, been little empirical examination of teacher training as part of a quality-development system for learning and instruction. The concept of learning effectiveness is barely touched upon in teacher training. Tellingly, inservice training gets only a peripheral mention in articles on instructional quality: The QAIT and MACRO models are prime examples here (see Ditton, 2000, for an overview). QAIT covers the main dimensions of instruction: Quality (clarity, structure, comprehensibility), Appropriateness (difficulty, pace, diagnostic competence), Incentives to learn (interest) and Time (time management, class management). MACRO covers characteristics of the school environment: Meaningful, universally understood goals, Attention to daily academic functioning; Coordination among programs and between school and parents over time; Recruitment of teachers and development of all staff; and Organization of the school to support universal student learning. Although instructional quality is often defined in terms of variables that are relatively easy to measure (Clausen, 2000), its relationship to teacher behavior and student performance is not always established. Criblez (2001), on the other hand, states that, "The outcome of effective teacher training is effective instruction provided by a qualified teacher" (p. 109, own translation). To realize this concept of effectiveness, both teacher variables (professional knowledge, teacher behavior) and student variables (student behavior, performance, motivation) must be controlled. Because it is not empirically possible to determine in retrospect whether teachers acquired their pedagogical skills in or after training, any investigation of the causes for current teacher behavior that does not relate specifically to a given training module will have a subjective component. If the goals of teacher training were formulated as standards in the same way as the goals of school education, however, the effectiveness of programs targeting these goals could be controlled systematically by assessing observable indicators of teacher behavior and student performance.

Accordingly, we must start by describing the professional knowledge of science teachers and determining which (successful or unsuccessful) profiles of basis models are needed to supplement this knowledge. On this basis, and taking the abovementioned concepts for teacher training into account, a learning-process oriented module for teacher training can be devised, and its effects on teacher and student behaviors and student performance can be examined and controlled.

Summary

Research aiming to improve the quality of school instruction should cover at least the following five elementary domains: (a) the teaching profession, (b) student performance, (c) efforts to structure instruction using the findings and insights of relevant research, (d) the system and structure of the individual school, and (e) individual learning processes. Of course, research efforts tend to focus on the content and structure of specific subjects and the way these are learned. However, the effects on student performance and behavior of variations in lesson content have rarely been evaluated, and the few available studies comparing traditional lessons and new approaches (e.g., Starauschek, 2001, on the Karlsruhe physics course, a nontraditional approach to physics education; Weber, 2002, on a new way of presenting optics using Fermat's principle) have found weak, if any, effects. It can be assumed that the factors outlined in the present article override any effects of purely content-driven changes to science instruction. Reyer (2004) described how lower secondary students responded to the intentions of a physics teacher and established that in 8th- and 9th-grade physics lessons student performance did not correlate with the teachers' intentions. The student deficits observed there corresponded with the PISA deficits and could be attributed to the fact that the students were not given adequate opportunity to engage in the intended learning processes (e.g., problem solving or conceptual change).

An exclusively content-driven approach, viewed as educational reduction, means that teaching is stripped of any academic basis. Academic standards for teaching the scientific disciplines cannot be derived from the sciences themselves: (a) The research topics of science education do not correspond to those of the sciences, and (b) research on science education does not involve biological, chemical, or physical research. However, research on science education rarely meets the standards of empirical social research either. This means that instruction and initial and inservice teacher training are developed in a rather intuitive manner and that their outcomes are difficult to predict.

On the other hand, the fact that instructional research in the fields of pedagogy and educational psychology is not oriented toward subject-specific structures and competencies produces general insights into the quality of teaching that are, as a rule, empirically sound. Nevertheless, the criteria developed have rarely been applied and adapted in the classroom. Likewise, the effects of corresponding training programs on teacher behavior and student performance have rarely been tested. This is reflected by the scarcity of articles on this kind of instructional development in peer-reviewed national or international journals and by the modest amount of project-linked research funding allocated to such endeavors (Leutner, 2000).

These findings assume alarming proportions considering the need for research and development on science education that is already becoming apparent in the short term and will certainly be needed in the medium term in conjunction with planned quality-assurance measures. It will be necessary to develop assessment and control instruments, inservice training programs, and consulting strategies for schools that meet the standards of empirical research and are geared at specific subjects. This can only be achieved if didactics specialists work hand in hand with psychologists and educationalists. As the task of developing instruments of this kind for German, foreign languages, and mathematics instruction at various grade levels is already imminent, the science subjects will follow.

Without wanting to overrate the studies, TIMSS and PISA have certainly brought a new perspective to instructional research. As mentioned above, however, neither study allows the relations between student performance and instruction or teacher behavior (in specific subjects) to be properly examined. In our opinion, this field of research can only be addressed in a collaborative effort involving representatives of the relevant disciplines: science education, educational research, and the psychology of instruction and learning.

References

American Association for the Advancement of Science. (1989). *Project 2061: Science for all Americans*. Washington, DC: Author.

Artelt, C. (2000). *Strategisches Lernen* [Strategically learning]. Münster, Germany: Waxmann.

Ausubel, D. P. (1963). *The psychology of meaningful verbal learning*. New York: Grune & Stratton.

Bates, G. R. (1978). The role of laboratory in secondary school science programs. In M. B. Browe (Ed.), *What research says to the science teacher* (pp. 55–82). Washington, DC: National Science Teachers' Association.

Bauer, K.-O. (1997). Pädagogische Professionalität und Lehrerarbeit [Pedagogical professionality and teachers' work]. *Pädagogik*, *49*(4), 22–26.

Bauer, K.-O., Kopka, A., & Brindt, S. (1999). *Pädagogische Professionalität und Lehrerarbeit. Eine qualitativ empirische Studie über professionelles Handeln und Bewusstsein* [A qualitative-empirical study about professional behavior and consciousness]. Weinheim/München, Germany: Juventa.

Baumert, J. (1993). Lernstrategien, motivationale Orientierung und Selbstwirksamkeits–überzeugungen im Kontext schulischen Lernens [Strategies of learning, motivational orientation, and self-esteem in the context of learning at schools]. *Unterrichtswissenschaft*, *21*, 327–354.

Baumert, J., & Artelt, C. (2002). Bereichsübergreifende Perspektiven [General perspective]. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, K.-J. Tillmann, & M. Weiß (Eds.), *PISA 2000—Die Länder der Bundesrepublik Deutschland im Vergleich* (pp. 219–235). Opladen, Germany: Leske & Budrich.

Baumert, J., Artelt, C., Carstensen, C. H., Sibberns, H., & Stanat, P. (2002). Untersuchungs–gegenstand, Fragestellungen und technische Grundlagen der Studie [Topic, questions, and technical basis of the PISA study]. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, K.-J. Tillmann, & M. Weiß(Eds.), *PISA 2000—Die Länder der Bundesrepublik Deutschland im Vergleich* (pp. 11–38). Opladen, Germany: Leske & Budrich.

Baumert, J., Artelt, C., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., et al. (Eds.). (2002). *PISA 2000: Die Länder der Bundesrepublik Deutschland im Vergleich* [PISA 2000: Comparing the German states]. Opladen, Germany: Leske & Budrich.

Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., et al. (Eds.). (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* [International comparison of students' basic competences]. Opladen, Germany: Leske & Budrich.

Baumert, J., & Köller, O. (1996). Lernstrategien und schulische Leistungen [Learning strategies and performance at school]. In J. Möller, & O. Köller (Eds.), *Emotionen, Kognitionen und Schulleistung* (pp. 137–154). Weinheim, Germany: Beltz.

Baumert, J., & Köller, O. (2000). Unterrichtsgestaltung, verständnisvolles Lernen und multiple Zielerreichung im Mathematik- und Physikunterricht der gymnasialen Oberstufe [Arrangement of lessons, meaningful learning, and achievement of multiple aims in mathematics and physics lessons of upper secondary schools]. In J. Baumert, W. Bos, & R. Lehmann (Eds.), *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie-Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn: Vol. 2. Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe* (pp. 271–315). Opladen, Germany: Leske & Budrich.

Baumert, J., & Lehmann, R. (1997). *TIMSS—Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich* [International comparison of mathematics and science education]. Opladen, Germany: Leske & Budrich.

Baumert, J., Stanat, P., & Demmrich, A. (2001). PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung in der Schule [Topic, questions, and technical basis of the PISA study]. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, K.-J. Tillmann, & M. Weiß(Eds.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (pp. 15–68). Opladen, Germany: Leske & Budrich.

Baumgartner, P., & Payr, S. (1994). *Lernen mit Software* [Learn with software]. Innsbruck, Austria: Österreichischer Studienverlag.

Beaton, A. E., Martin, M. O., Mullis, I. V. S., Gonzales, E. J., Smith, T. A., & Kelly, D. L. (1996). *Science achievement in the middle school years: IEA's* *third international mathematics and science study*. Chestnut Hill, MA: Boston College.

Bellenberg, G., Böttcher, W., & Klemm, K. (2001). *Stärkung der Einzelschule* [Stabilization of the single school]. Neuwied, Germany: Luchterhand.

Bingle, W. H., & Gaskell, P. J. (1994). Scientific literacy for decision making and the social construction of scientific knowledge. *Science Education*, *78*, 185–201.

Biological Science Curriculum Study. (1990, March). BSCS and educational reform. *BSCS: The Natural Selection*, *1*.

Blankertz, H. (1985). *Berufsbildung und Utilitarismus: Problemgeschichtliche Untersuchungen*. [Professional development and utilitarianism: Analysis of the history of problems]. Weinheim, Germany: Juventa.

Borchert, J., Dahbashi, A., & Knopf-Jerchow, H. (Eds.). (1996). Das Konzept eines innerschulischen Lehrertrainings unter Einbeziehung verbaler und videogestützter Rückmeldungen [Concept of in service teacher training using video feedback]. In W. Pallasch, W. Mutzek, & H. Reimers (Eds.), *Beratung—Trainin—Supervision. Eine Bestandsaufnahme über Konzepte zum Erwerb von Handlungskompetenz in pädagogischen Arbeitsfeldern* (pp. 53–62). Weinheim, Germany: Juventa.

Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Walther, G., & Valtin, R. (2003). *Erste Ergebnisse ais IGLU* [First results of IGLU]. Münster, Germany: Waxmann.

Brouër, B. (2001). *Förderung der Wahrnehmung von Lernprozessen durch die Anwendung der Basismodelle des Lernens bei der Gestaltung von Unterricht* [Supporting the perception of learning processes using basis models of learning]. *Unterrichtswissenschaft*, *29*, 153–170.

Bulmahn, E., Wolff, K., & Klieme, E. (2003). *Zur Entwicklung nationaler Bildungsstandards: Eine Expertise* [The development of national standards of education: An expertise]. Berlin, Germany: BMWF.

Bund-Länder-Kommission-Expertise. (1997). Gutachten zur Vorbereitung des Programms "Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts" [Expertise of the program "Improvement of the Efficiency of Mathematics and Science Lessons"]. Bonn, Germany: Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung.

Bybee, R. W. (1997). Toward an understanding of scientific literacy. In W. Gräber & C. Bolte (Eds.), *Scientific literacy* (pp. 37–68). Kiel, Germany: IPN.

Champagne, A. B., Gunstone, R. F., & Klopfer, L. E. (1985). Instructional consequences of students' knowledge about physical phenomena. In L. H. T. West & A. L. Pines (Eds.), *Cognitive structure and conceptual change* (pp. 163–188). New York: Academic.

Chomsky, N. (1965). Aspects of the theory of syntax. Cambridge, MA: MIT Press.

Clausen, M. (2000). *Wahrnehmung von Unterricht/Übereinstimmung, Konstruktvalidität, und Kriteriumsvalidität in der Forschung zur Unterrichtsqualität* [Perception of teaching and learning/correspondence, construct validity, and criterion validity regarding research on quality of instruction]. Dissertation am Fachbereich Erziehungswissenschaft und Psychologie, Freie Universität, Berlin, Germany.

Criblez, L. (2001). Die Wirksamkeit der Lehrerbildungssysteme in der Schweiz [The effectiveness of teacher-training systems in Switzerland]. In F. Oser & J. Ölkers (Eds.), *Die wirksamkeit der lehrerbildungssysteme* (pp. 99–140). Zürich, Switzerland: Rüegger.

DeBoer, G. E., & Bybee, R. W. (1995). The goals of science curriculum. In R. W. Bybee, & J. D. McInerney (Eds.), *Redesigning the science curriculum: A report on the implications of standards and benchmarks for science education* (pp. 71–74). Colorado Springs, CO: BSCS.

Dewe, B. (2000). *Betriebspädagogik und berufliche Weiterbildung: Wissenschaf—Forschung—Reflexion* [Professional education and professional training: Investigation and reflection]. Bad Heilbrunn, Germany: Klinkhardt.

Diederich, J., & Tenorth, H.-E. (1997). *Theorie der Schule. Ein Studienbuch zur Geschichte, Funktionen und Gestaltung* [Theory of school: A textbook regarding history, function, and structuring]. Berlin, Germany: Cornelsen Scriptor.

Ditton, H. (2000). Qualitätskontrolle und Qualitätssicherung in Schule und Unterricht: Eine Übersicht über den empirischen Forschungsstand [Quality control and management at schools and in classrooms: An overview of recent empirical investigations]. *Zeitschrift für Pädagogik*, *41*(Suppl.), 73–92.

Dörner, D. (1996). Lernmotivation [Learning motivation]. In J. Hoffmann & W. Kintsch (Eds.), *Enzyklopädie der Psychologie, Themenbereich C Theorie und Forschung, Ser. 2 Kognitionen, Bd. 7 Lernen* (pp. 179–202). Göttingen, Germany: Hogrefe.

Duit, R., Fischer, H. E., & Müller, W. (2002). Vielfalt und Anregung statt Routine: Der Physikunterricht—braucht eine andere Aufgabenkultur [Multitude and effort instead of routine: Physics lessons need a new type of tasks]. *Unterricht Physik*, *67*, 4–7.

Dunbar, K., & Klahr, D. (1989). Developmental differences in scientific discovery processes. In D. Klahr, & K. Kotovsky (Eds.), *Complex information processing* (pp. 109–143). Hillsdale, NJ: Erlbaum.

Euler, D. (1994). Multi-mediales lernen: Theoretische Fundierungen und Forschungsstand [Multi-media learning: Theoretical basis and state of research]. *Unterrichtswissenschaft*, *22*, 291–311.

Eylon, B., & Linn, M. C. (1988). Learning and instruction: An examination of four research perspectives in science education. *Review of Educational Research*, *58*, 251–301.

Fend, H. (2001). *Qualität im Bildungswesen: Schulforschung zu Systembedingungen, Schulprofilen, und Lehrerleistung* [Quality of the education system: Research on school system, school profiles and teacher performance]. Weinheim, Germany: Juventa.

Fischer, H. E. (1989). *Lernprozesse im Physikunterricht: Falluntersuchungen im Unterricht zur Elektrostatik aus konstruktivistischer Sicht* [Learning processes in physics lessons: Case studies of electrostatic lessons from a constructivist perspective]. Dissertation, Universität Bremen, Fachbereich, Germany.

Fischer, H. E. (1993). Framework for conducting empirical investigations of learning processes. *Science Education*, *77*, 131–152.

Fischer, H. E. (1998). Scientific literacy und Physiklernen [Scientific literacy and learning physics]. *Zeitschrift für Didaktik der Naturwissenschaften*, *4*(2), 41–52.

Fischer, H. E. (1999). Ein handlungs und kommunikationsorientierter Einstieg in die Elektrostatik [Action and communication orientated introduction of electrostatics]. *NiU-Physik*, *10*(50), 16–20.

Fischer, H. E. (2001). Einfuhrüng in die Konstruktion und Bewertung von Physikaufgaben zur Verbesserung des Physikunterrichts [Introduction to construction and assessment of physics tasks]. In Ch. Burkhard, H.-H. Weiß, & D. Meyer (Eds.), *Qualitätsentwicklung und Qualitätssicherung von Unterricht in der Sekundarstufe I* (pp. 37–53). Soest, Germany: Landesinstitut für Schule und Weiterbildung.

Fischer, H. E., & Breuer, E. (1997). Elektrostatik im Physik-Leistungskurs: Ein spiel- und kommunikationsorientierter Einstieg [Electrostatics in upper secondary classes. Action and communication orientated introduction]. In H. E. Fischer (Ed.), *Handlungsorientierter Physik-Unterricht Sekundarstufe II* (pp. 6–28). Bonn, Germany: Dümmler.

Fischer, H. E., & Draxler, D. (2001). Aufgaben und naturwissenschaftlicher Unterricht [Tasks and scientific instruction]. *Der Mathematische und Naturwissenschaftliche Unterricht*, *54*, 388–393.

Fischer, H. E., & Draxler, D. (2002). Unterrichtspraxis: Konstruktion und Bewertung von Physikaufgaben [Teaching: Construction and assessment of physics tasks]. In E. Kircher (Ed.), *Physikdidaktik in der Praxis* (pp. 300–322). Berlin, Germany: Springer.

Fischer, H. E., & Horstendahl, M. (1997). Motivation and learning physics [in special issue about European research in science education]. *Research in Science Education*, *27*, 411–424.

Fischer, H. E., Reyer, T., Wirz, T., Bos, W., & Höllrich, N. (2002). Unterrichtsgestaltung und Lernerfolg im Physikunterricht [The arrangement and learning outcomes of physics lessons]. *Zeitschrift für Pädagogik*, *45*(Suppl.), 124–138.

Fischler, H., & Zedler, P. (1999). *Unterrichtsskripts und lehrerexpertise: Bedingungen ihrer Modifikation. Erstantrag zum DFG-Schwerpunktprogramm BIQUA* [Lesson scripts and teacher expertise: Requirements of modifications]. Berlin, Germany: Erfurt.

Frey, K., & Achtenhagen, F. (1975). *Curriculum-Handbuch Bd. 1–3* [Handbook of curriculum development (Vol. 1–3)]. *IPN-Kiel an der Universität Kiel*. München, Germany: Piper.

Fuhrmann, M. (1999). *Der Europäische Bildungskanon des Bürgerlichen Zeitalters* [The European canon of general education during the bourgeois age]. Frankfurt/ Leipzig, Germany: Insel.

Gabel, D. L., & Bunce, D. M. (1994). Research and problem solving: Chemistry. In D. L. Gabel (Ed.), *Handbook of research on science teaching and learning. A project of the National Science Teacher Association* (pp. 301–326). New York: Macmillan.

Gallagher, J. J., & Tobin, K. (1987). Teacher management and student engagement in high school science. *Science Education*, *71*, 535–555.

Glasersfeld, E. V. (1995). *Radical constructivism: A way of knowing and learning*. London: Falmer.

Glynn, S. M., Muth, K., & Denise (1994). Reading and writing to learn science: Achieving scientific literacy. *Journal of Research in Science Teaching*, *31*, 1057–1074.

Gräber, W., & Bolte, C. (Eds.). (1997). *Scientific literacy*. Kiel, Germany: IPN.

Gräber, W., Nentwig, P., Koballa, T., & Evans, R. (2002). *Scientific literacy*. Opladen, Germany: Leske & Budrich.

Guillon, A. (1995). Démarches scientifiques en travaux pratiques de physique de DEUG à l'Université Cergy Pontoise [Students' development of scientific methods during physics experiments at University of Cergy Pontoise]. *Didaskalia*, *7*, 113–127.

Haller, K. (1999). Über den Zusammenhang von Handlung und Zielen: Eine empirische Untersuchung zu Lernprozessen im physikalischen Prakticum [The correlation between action and aims: Empirical investigation of learning processes during experimental training of physics]. Berlin, Germany: Logos.

Harland, T. (2002). Zoology students' experiences of collaborative enquiry in problem-based learning. *Teaching in Higher Education*, *7*(1), 3–15.

Häußler, P., & Lind, G. (1998). *BLK-Programmförderung "Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts". Erläuterungen zu Modul 1 mit Beispielen für den Physikunterricht. Weiterentwicklung der Aufgabenkultur im mathematisch-naturwissenschaftlichen Unterricht* [BLK-development program "Improvement of the efficiency of mathematics and science lessons" Explication of Module 1 including examples for physics lessons. Development of tasks for mathematics and science lessons]. Kiel, Germany: IPN.

Helmke, A. (2000). TIMSS und die Folgen: Der weite Weg von der externen Evaluation der Schulleistungen zur Verbesserung des Lehrens und Lernens [TIMSS and the consequences: The long way from external evaluation of students' performance to better teaching and learning]. In U. P. Trier (Ed.), *Bildungswirksamkeit zwischen Forschung und Politik* (pp. 135–164). Bern and Chur, Switzerland: Rüegger.

Helmke, A., & Weinert, F. E. (1997). Bedingungsfaktoren schulischer leistung [Condition factors of school achievement]. In F. E. Weinert (Ed.), *Psychologie des Unterrichts und der Schule* (pp. 71–176). Göttingen, Germany: Hogrefe.

Heymann, H. W. (1994). Über Mathematikunterricht unter dem anspruch von Allgemeinbildung [Over mathematics instruction under the requirement of general education]. In M. A. Meyer & W. Plöger (Eds.), *Allgemeine didaktik: Fachdidaktik und fachunterricht*. Weinheim, Germany: Beltz.

Heymann, H. W. (1996). *Allgemeinbildung und Mathematik* [General education and mathematics]. Weinheim, Germany: Beltz.

Hofstein, A., & Lunetta, V. N. (1982). The role of the laboratory in science teaching: Neglected aspects for research. *Review of Educational Research*, *52*, 201–217.

Horstendahl, M., Fischer, H. E., & Rolf, R. (1999). Konzeptuelle und motivationale Aspekte der Handlungsregulation von Schülerinnen und Schülern im Experimentalunterricht der Physik [Conceptual and motivational aspects of the action regularization of pupils and pupils in the experimental instruction of physics]. *Zeitschrift für Didaktik der Naturwissenschaften*, *9*, 7–26.

Horster, L., & Rolff, H.-G. (2001). *Unterrichtsentwicklung, Grundlagen, Praxis, Prozesssteuerung* [Development of instruction, requirements, practice]. Weinheim, Germany: Beltz.

Hucke, L. (2000). *Handlungsregulation und Wissenserwerb in traditionellen und computergestützten Experimenten des physikalischen Praktikums* [Action regulation and assessment in traditional and computer-aided experiments of university physics courses]. Berlin, Germany: Logos.

Hucke, L., & Fischer, H. E. (2001). Fachdidaktische Forschung zur Verbesserung der experimentellen naturwissenschaftlichen Ausbildung: Eine Untersuchung im physikalischen Anfängerpraktikum [Research on subject education to increase the experimental instruction of science education at university]. In C. Finkbeiner & G. W. Schnaitmann (Eds.), *Lehren und Lernen im Kontext empirischer Forschung und Fachdidaktik* (pp. 496–517). Donauwörth, Germany: Ludwig Auer.

Hucke, L., & Fischer, H. E. (2002). The link of theory and practice in traditional and in computer-based university laboratory experiments. In D. Psillos & H. Niedderer (Eds.), *Teaching and learning in the science laboratory: A book based on the European project "Labwork in Science Education"* (pp. 205–218). Dordrecht, The Netherlands: Kluwer Academic.

Hussy, W. (1984). *Denkpsychologie: Ein Lehrbuch (Bd. 1)* [Cognitive psychology. A textbook (Vol. 1)]. Stuttgart, Germany: Kohlhammer.

Jacobs, J. K., Kawanaka, T., & Stigler, J. W. (1999). Integrating qualitative and quantitative approaches to the analysis of video data on classroom teaching. *International Journal of Educational Research*, *31*, 717–724.

Jung, W. (1970). *Beiträge zur Didaktik der Physik* [Contributions to physics education]. Frankfurt, Germany: Diesterweg.

Kempfert, G., & Rolff, H.-G. (1999). *Pädagogische Qualitätsentwicklung* [Pedagogical quality development]. Weinheim, Germany: München.

Kennedy, M. F. (1994, March). Instructional design or personal heuristics in classroom instructional planning. *Educational Technology*, *34*(4), 17–25.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, *12*, 1–48.

Klahr, D., Dunbar, K., & Fay, A. L. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology*, *25*, 111–146.

Klauer, K. J. (1985). Framework for a theory of teaching. *Teaching and Teacher Education*, *1*, 5–17.

Klemm, K. (1998). Steuerung der Schulentwicklung durch zentrale Leistungskontrollen [Control of school development using central assessment]. In H.-G. Rolff, K.-O. Bauer, K. Klemm, & H. Pfeiffer (Eds.), *Jahrbuch der Schulentwicklung 10* (pp. 271–294). Weinheim, Germany: Juventa.

Klieme, E. (2000). Fachleistungen im voruniversitären Mathematik- und Physikunterricht: Theoretische Grundlagen, Kompetenzstufen and Unterrichtsschwerpunkte [Students' achievement as a result of preuniversity mathematics and physics lessons]. In J. Baumert, W. Bos, & R. Lehmann (Eds.), *Dritte internationale Mathematik-und Naturwissenschaftsstudie—Mathematische und naturwissenschaftliche Bildung am End der Schullaufbahn* (Vol. 2, pp. 57–128). Leverkusen, Germany: Leske & Budrich.

Klieme, E. (2001). Fachleistungen im voruniversitären Mathematik- und Physikunterricht: theoretische und methodische Grundlagen [Performance of preuniversity mathematics and science instruction: Theoretical and methodological basis]. In J. Baumert, W. Bos, & R. Lehmann (Eds.), *TIMSS/III: Schülerleistungen in Mathematik und den Naturwissenschaften am Ende der Sekundarstufe* (Vol. 2, pp. 57–128). Opladen, Germany: Leske und Budrich.

Klieme, E. (2003). Bildungsstandards. Ihr Beitrag zur Qualitätsentwicklung im Schulsystem [Standards of education: Their contribution to quality development of the school system]. *Die Deutsche Schule*, *95*(1), 10–16.

Klieme, E., & Baumert, J. (2001). Mathematisch-naturwissenschaftliche Grundbildung: Konzeptuelle Grundlagen und die Erfassung und Skalierung von Kompetenzen [Mathematical and science education: Conceptual basis and the assessment of competences]. In J. Baumert, W. Bos, & R. Lehmann (Eds.), *TIMSS/III: Schülerleistungen in Mathematik und den Naturwissenschaften am Ende der Sekundarstufe II* (Vol. 1, pp. 85–134). Opladen, Germany: Leske und Budrich.

Klieme, E., & Bos, W. (2000). Mathematikleistung und mathematischer Unterricht in Deutschland und Japan. Triangulation qualitativer und quantitativer Analysen am Beispiel der TIMS-Studie [Performance and description of mathematic lessons in Germany and Japan: Triangulation of qualitative and quantitative data of the TIMS-study]. *Zeitschrift für Erziehungswissenschaft*, *3*, 359–379.

Klieme, E., Funke, J., Leutner, D., Reimann, P., & Wirth, J. (2001). Problemlösen als fächerübergreifende Kompetenz? Konzeption und erste Resultate aus einer Schulleistungsstudie [Problem solving as general competence? Design and first results of performance assessment at schools]. *Zeitschrift für Pädagogik*, *47*, 179–200.

Klieme, E., & Schecker, H. (2001). Mehr denken, weniger rechnen [More thinking, less calculating]. *Physikalische Blätter*, *57*(7/8), 113–117.

Klippert, H. (2000). *Pädagogische Schulentwicklun* [Pedagogical school development]. Weinheim, Germany: Juventa.

Korte, H. (1998). Einführung in die Geschichte der Soziologie [Introduction of history of sociology]. Opladen, Germany: Leske & Buderich.

Krapp, A. (1993). Die Psychologie der Lernmotivation [The psychology of the learning motivation]. *Zeitschrift für Pädagogik*, *39*, 187–206.

Labudde, P. (1993). *Erlebniswelt physik* [Experience world physics]. Bonn, Germany: Dümmler.

Labudde, P. (2000). *Konstruktivismus im Physikunterricht der Sekundarstufe II* [Constructivism and physics instruction in upper secondary level]. Wien, Germany: Haupt.

Labudde, P., & Pfluger, D. (1999). Physikunterricht in der Sekundarstufe II: Eine empirische Analyse der Lern-Lehr-Kultur aus konstruktivistischer Perspektive [Physics instruction in the upper secondary level: An empirical analysis of learning and reading culture from a constructivist perspective]. *Zeitschrift für Didaktik der Naturwissenschaften*, *5*(2), 33–50.

Lange, H. (1999). Qualitätssicherung in Schulen [Quality assurance in schools]. *Die Deutsche Schule*, *2*, 144–159.

Ledermann, L. (2001). Revolution in science education: Put physics first! *Physics Today*, *54*(9), 11–12.

Lehrplan Naturwissenschaften (Physik, Chemie, Biologie) Gesamtschule. (1999). In Ministerium für Schule und Weiterbildung, Wissenschaft und Forschung des Landes Nordrhein-Westfalen (Ed.), *Psychology of the motivation to learn*. Frechen, Germany: Ritterbach.

Leopold, C., & Leutner, D. (2002). Der Einsatz von Lernstrategien in einer konkreten Lernsituation bei Schülern unterschiedlicher Jahrgangsstufen [Students of different age using learning strategies in concrete situations]. *Zeitschrift für Pädagogik*, *45*(Suppl.), 240–258.

Leutner, D. (1992). *Adaptive Lehrsysteme: Instruktionspsychologische Grundlagen und experimentelle Analysen* [Adaptive teaching systems: Requirements of instructional psychology and experimental analysis]. Weinheim, Germany: Beltz–Psychologie Verlags Union.

Leutner, D. (1997). Instruktions-Design im Spannungsfeld von Behaviorismus, Kognitionspsychologie und Konstruktivismus [Instructional design in the area of conflict of behaviorism, cognitive psychology, and constructivism]. In H. Bayrhuber, U. Gebhard, K.-H. Gehlhaar, D. Graf, H. Gropengießer, U. Harms, U. Kattmann, R. Klee, & J. C. Schletter (Eds.), *Biologieunterricht und Lebenswirklichkeit* (pp. 79–97). Kiel, Germany: IPN.

Leutner, D. (1998a). Instruktionspsychologie [Dictionary of pedagogical psychology]. In D. H. Rost (Ed.), *Handwörterbuch Pädagogische Psychologie* (pp. 198–205). Weinheim, Germany: Psychologie Verlags Union.

Leutner, D. (1998b). Programmierter und computerunterstützter Unterricht [PC-supported instruction]. In D. H. Rost (Ed.), *Handwörterbuch Pädagogische Psychologie* (pp. 404–409). Weinheim, Germany: Psychologie Verlags Union.

Leutner, D. (2000). Blick zurück nach vorn: Trends der Lehr-Lernforschung [Looking back on the future: Trend of research on teaching and learning]. *Unterrichtswissenschaft*, *28*, 30–37.

Leutner, D. (2002). The fuzzy relationship of intelligence and problem solving in computer simulations. *Computers in Human Behavior*, *18*, 685–697.

Leutner, D. (2004). Adaptivity in open learning environments. In N. M. Seel & S. Dijkstra (Eds.), *Curriculum, plans and processes of instructional design: International perspectives* (pp. 289–307). Mahwah, NJ: Erlbaum.

Leutner, D., & Leopold, C. (2003a). Lehr-lernpsychologische Grundlagen selbstregulierten lernens [Psychological basis of self-regulated learning]. In U. Witthaus, W. Wittwer, & C. Espe (Eds.), *Selbst gesteuertes Lernen—Theoretische und praktische Zugänge* (pp. 43–67). Bielefeld, Germany: Bertelsmann.

Leutner, D., & Leopold, C. (2003b). Selbstreguliertes Lernen als Selbstregulation von Lernstrategien. Ein Trainingsexperiment mit Berufstätigen zum lerNen mit Sachtexten [Self-regulated learning as self-regulation of learning strategies]. *Unterrichtswissenschaft*, *31*, 38–56.

Lompscher, J. (1994). *Lernstrategien: Zugänge auf der Reflexions- und Handlungsebene, LLF-Berichte* [Strategies of learning: The perspective of reflection and action] (Vol. 9, pp. 114–129). Potsdam, Germany: Universität Potsdam.

Lunetta, V. N. (1998). The school science laboratory: Historical perspectives and contexts for contemporary teaching. In K. Tobin & B. Fraser (Eds.), *International handbook of science education* (pp. 249–264). Amsterdam: Kluwer.

Mandl, H., Gruber, H., & Renkl, A. (1995). Situiertes Lernen in multimedialen Lernumgebungen [Situated learning in multimedia learning environments]. In L. J. Issing & P. Klimsa (Eds.), *Information und Lernen mit Multimedia* (pp. 167–178). Weinheim, Germany: Psychologie Verlags Union.

Mandl, H., & Hron, A. (1989). Psychologische Aspekte des Lernens mit dem Computer [Psychological aspects of learning using computers]. *Zeitschrift für Pädagogik*, *35*, 657–678.

Mayer, R. E. (2001). *Multimediales Lernen* [Multimedial learning]. Cambridge, England: Cambridge University Press.

Merrill, M. D. (1991). Constructivism and instructional design. *Educational Technology*, *31*, 45–53.

Millar, R. (1993). Physics in Salters' Science. *Physics Education*, *28*, 356–361.

Millar, R., & Driver, R. (1987). Beyond processes. *Studies in Science Education*, *14*, 33–62.

Moallem, M., & Earle, R. S. (1998). Instructional design models and teacher thinking: Toward a new conceptual model for research and development. *Educational Technology*, *38*(2), 5–22.

Möller, K., Tenberge, C., & Ziemann, U. (1996). *Technische Bildung im Sachunterricht: Eine quantitative Studie zum Sachunterricht an Nordrhein-Westfälischen Grundschulen. Veröffentlichungen der Abteilung Didaktik des Sachunterrichts/ Institut für Forschung und Lehre der Primarstufe* (Vol. 2) [Technical education in primary school: A quantitative study at primary schools of North Rhine Westphalia]. Münster, Germany: Universität Münster.

MSWF—NRW. (2001). *Allgemeinbildende schulen. Statistische Übersicht Nr. 321* [Schools for general education]. Düsseldorf, Germany: Ministry of Education in NRW.

Mueller, C. M., & Dweck, C. S. (1998). Intelligence praise can undermine motivation and performance. *Journal of Personality and Social Psychology*, *75*, 33–52.

Mullis, I. V. S., Martin, M. O., & Gonzalez, E. J. (2004). *International achievement in the processes of reading comprehension: Results from PIRLS 2001 in 35 countries*. Chestnut Hill, MA: Boston College.

National Science Teacher Association. (1990). *Science education initiatives for the 1990s*. Washington, DC: Author.

Organization for Economic Cooperation and Development (OECD). (2001). *Knowledge and skills for life: First results from the OECD Programme for International Student Assessment (PISA) 2000*. Paris: Author.

Organization for Economic Cooperation and Development (OECD). (2003). *The PISA 2003 assessment framework : Mathematics, reading, science and problem solving knowledge and skills*. Paris: Author.

Oser, F. K., & Baeriswyl, F. J. (2001). Choreographies of teaching: Bridging instruction to learning. In V. Richardson (Ed.), *AERA's handbook of research on teaching* (4th ed., pp. 1031–1065). Washington, DC: American Educational Research Association.

Oser, F. K., & Patry, J.-L. (1994). Sichtstruktur und Basismodelle des Unterrichts: Über den Zusammenhang von Lehren und Lernen unter dem Gesichtspunkt psychologischer Lernverläufe [Apparent structure and basis models of lessons: The correlation of teaching and learning regarding learning processes]. In R. Olechowski & B. Rollett (Eds.), *Theorie und Praxis—Aspekte empirisch-pädagogischer Forschung— quantitative und qualitative Methoden* (pp. 138–146). Frankfurt, Germany: Lang.

Oser, F. K., Patry, J.-L., Elsässer, T., Sarasin, S., & Wagner, B. (1997). *Choreographien unterrichtlichen Lernens: Schlußbericht an den Schweizerischen Nationalfonds zur Förderung der wissenschaftlichen Forschung* [Choreographies of learning at schools: Final report for the Swiss National Science Foundation]. Freiburg, Switzerland: Pädagogisches Institut der Universität Freiburg.

Pallasch, W. (1997). *Supervision: Neue Formen beruflicher Praxisbegleitung in pädagogischen arbeitsfeldern* [Supervision: New types of professional supervising of pedagogical practice]. Weinheim/München, Germany: Juventa.

Pallasch, W., Mutzek, W., & Reimers, H. (1992). *Beratung—Training—Supervision. Eine Bestandsaufnahme über Konzepte zum Erwerb von Handlungskompetenz in pädagogischen Arbeitsfeldern* [Advice, training, and supervision:. Stocktaking of concepts to acquire competence for pedagogical practice]. Weinheim, Germany: Juventa.

Parchmann, I., Demuth, R., Ralle, B., Paschmann, A., Huntemann, H., Krilla, B., et al. (2001). Chemie im Kontext: Begründung und Realisierung eines Lernens in sinnstiftenden Kontexten [Chemistry in a context: Foundation and realization of learning in meaningful contexts]. *Praxis der Naturwissenschaften/Chemie*, *50*(1), 2–15.

Prenzel, M., Häußler, P., Rost, J., & Senkbeil, M. (2002). Der PISA Naturwissenschaftstest: Lassen sich die Aufgabenschwierigkeiten vorhersagen? [The PISA science test: Is it possible to predict the difficulty of tasks?] *Unterrichtswissenschaft*, *30*, 120–135.

Prenzel, M., Rost, J., Senkbeil, M., Häußler, P., & Klopp, A. (2001). Naturwissenschaftliche Grundbildung: Testkonzeption und Ergebnisse [Science education: Test concept and results]. In J. Baumert et al. (Eds.), *PISA 2000* (pp. 191–248). Opladen, Germany: Leske & Budrich.

Preußler, W. (1997). Effekte des Kontexts auf den Wissenserwerb bei der Steuerung eines dynamischen Systems [The effect of contexts on knowledge acquisition during controlling a dynamical system]. *Sprache und Kognition*, *16*(1), 48–59.

Radtke, F.-O. (1996). *Wissen und Können: Die Rolle der Erziehungswissenschaft in der Erziehung* [Knowledge and ability: The role of pedagogy in education]. Opladen, Germany: Leske & Buderich.

Reyer, Th. (2004). *Oberflächenmerkmale und Tiefenstrukturen im Unterricht: Exemplarische analysen im Physikunterricht der gymnasialen Sekundarstufe I* [Apparent structure and deep structure of lessons: Exemplary investigation of physics lessons at secondary level of grammar schools]. Berlin, Germany: Logos.

Rolff, H.-G., Buhren, C. G., Lindau-Bank, D., & Müller, S. (1998). *Manual Schulentwicklung.* [Manual of school development]. Weinheim, Germany: Beltz.

Rosen, S. A. (1954). History of the physics laboratory in the American public schools. *American Journal of Physics*, *22*, 194–204.

Roth, W.-M. (1995). *Authentic school science: Knowing and learning in open-inquiry science laboratories*. Dordrecht, The Netherlands: Kluwer Academic.

Sander, F. (1999). *Handeln und Lernen im Praktikum* [Acting and learning in experimental training]. Dissertation am Fachbereich 1 (Physik/Elektrotechnik) der Universität Bremen, Bremen, Germany.

Schecker, H., & Winter, B. (Eds.). (1999). *Abschlussbericht des Modellversuchs BINGO* [Final report of the experiment BINGO]. Bremen, Germany: Senator für Bildung und Wissenschaft.

Schiefele, U., & Pekrun, R. (1996). Psychologische Modelle des fremdgesteuerten und selbstgesteuerten Lernens [Psychological models of acting under orders and self-determined]. In F. E. Weinert (Ed.), *Enzyklopädie der Psychologie. Themenbereich D, Praxisgebiete: Serie 1, Pädagogische Psychologie; Vol. 2, Psychologie des Lernens und der Instruktion* (pp. 249–278). Göttingen, Germany: Hogrefe.

Schreiber, B. (1998). *Selbstreguliertes Lernen* [Self-determined learning]. Münster, Germany: Waxmann.

Schwanitz, D. (1999). *Bildung*. Frankfurt, Germany: Eichborn.

Seel, N. M. (2000). *Psychologie des Lernens* [Psychology of learning]. Stuttgart, Germany: UTB.

Seidel, T., Prenzel, M., Duit, R., Euler, M., Geiser, H., Hoffmann, L., et al. (2002). "Jetzt bitte alle nach vorne schauen!":Lehr-Lernskripts im Physikunterricht und damit verbundene Bedingungen für individuelle Lernprozesse ["Look to the front please!" Teaching and learning scripts of physics lessons and related conditions of individual learning processes]. *Unterrichtswissenschaft*, *30*(1), 52–77.

Shamos, M. H. (1995). The myth of scientific literacy. *Liberal Education*, *82*(3), 44–49.

Simon, H. A., & Lea, G. (1974). Problem solving and rule induction: A unified view. In L. W. Gregg (Ed.), *Knowledge and cognition* (pp. 105–127). Hillsdale, NJ: Erlbaum.

Solomon, J., & Aikenhead, G. S. (1994). *STS education: International perspectives on reform*. New York: Teachers College.

Starauschek, E. (2001). *Physikunterricht nach dem Karlsruher Physikkurs: Ergebnisse einer Evaluationsstudie* [Physics lessons based on the "Karlsruher Physikkurs": Results of an evaluation]. Berlin, Germany: Logos-Verlag.

Stebler, R., Reusser, K., & Ramseier, E. (1998). Praktische Anwendungsaufgaben zur integrierten Förderung formaler und materialer Kompetenzen [Practical tasks for an integrated fostering of formal and material competences]. *Bildungsforschung und Bildungspraxis*, *20*(1), 28–54.

Stigler, J. W., Gonzales, P., Kawanaka, T., Knoll, S., & Serrano, A. (1999). *The TIMSS videotape classroom study: Methods and findings from an exploratory research project on eighth-grade mathematics instruction in Germany, Japan, and the United States.* Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Stigler, J. W., & Hiebert, J. (1997). Understanding and improving classroom mathematics instruction: An overview of the TIMSS Video Study. *Phi Delta Kappan*, *79*, 14–21.

Sumfleth, E. (1985). Lernhilfen für den problemlösenden Chemieunterricht [Learning tools for problem solving chemistry lessons]. *Chimica Didactica*, *11*, 63–88.

Sumfleth, E. (1988). *Lehr- und Lernprozesse im Chemieunterricht: Das Vorwissen des Schülers in einer kognitionspsychologisch fundierten Unterrichtskonzeption* [Teaching and learning processes during chemistry lessons: Preknowledge of students as a basis of a cognitive psychology based on a conception of teaching]. Frankfurt, Germany: Lang.

Sumfleth, E., Bergmann, D., & Dannat, P. (1990). Ein Unterrichtsvorschlag zum Thema Elektrochemie für die Sekundarstufe II [A proposal for teaching electrochemistry at upper secondary level of grammar schools]. *Der Mathematische und Naturwissenschaftliche Unterricht*, *43*(1), 31–40.

Sumfleth, E., & Dannat, P. (1988). Konzeption eines problemlösenden Chemieunterrichts für die gymnasiale Oberstufe [Concept of problem solving chemistry lessons at upper secondary level of grammar school]. *Deutscher Verein zur Förderung des mathematischen und naturwissenschaftlichen Unterrichts*, *41*, 25–40.

Sumfleth, E., & Pitton, A. (1998). Sprachliche Kommunikation im Chemieunterricht: Schülervorstellungen und ihre Bedeutung im Unterrichtsalltag [Communication in chemistry lessons: Students ideas and their relevance for everyday teaching]. *Zeitschrift für Didaktik der Naturwissenschaften*, *4(*2) 4–20.

Sumfleth, E., Ploschke, B., & Geisler, A. (1999). Schülervorstellungen und Unterrichtsgespräche zum Thema Säure-Base [Ideas of students and lesson talks about the topic acid-base]. In E. Sumfleth (Ed.), *Chemiedidaktik im Wandel* (pp. 91–115). Münster, Germany: Lit.

Sumfleth, E., & Stachelscheid, K. (1986). Carbonylverbindungen: Ein Unterrichtsvorschlag zur Einführung in die organische Chemie unter Einsatz der neuen Lernhilfen [Carbonyl combinations: A proposal to introduce organic chemistry using new learning tools]. *Chimica Didactica*, *12*, 137–148.

Sumfleth, E., Stachelscheid, K., & Gramm, A. (1989). Stoffe: Eigenschaften und Reaktionen–Modelle: Teilchenanordnungen und Umordnungen, eine mit Lernhilfen gestützte Einführung in die Chemie [Substances: Properties and reactions-models: Particle formation and rearrangement, learning aid supported introduction of chemistry]. *Der Mathematische und Naturwissenschaftliche Unterricht*, *42*, 411–422.

Sumfleth, E., Wild, E., Rumann, S., & Exeler, J. (2002). Wege zur Förderung der naturwissenschaftlichen Grundbildung deutscher Schülerinnen und Schüler im Chemieunterricht [Ways to support scientific literacy of German students in chemistry lessons]. *Zeitschrift für Pädagogik*, *45*(Suppl.), 207–221.

Tamir, P., & Lunetta, V. N. (1981). Inquiry related tasks in high school science laboratory handbooks. *Science Education*, *65*, 477–484.

Tennstädt, K.-C. (Ed.). (1992). *Das Konstanzer Trainingsmodell (KTM): Neue Wege im Schulalltag: Ein Selbsthilfeprogramm für zeitgemässes Unterrichten und Erziehen, Vol. 4* [The "Constance Training Model (CTM): New ways in everyday teaching: Self-help program of up-to-date teaching and educating]. Bern, Switzerland: Hans Huber.

Tenorth, H.-E. (1994). "Alles zu Lehren": Möglichkeiten und Perspektiven allgemeiner Bildung ["To teach everything": Possibilities and perspectives of general education]. Darmstadt, Germany: Wissenschaftliche Buchgesellschaft.

Theyßen, H. (2000). *Ein Physikpraktikum für Studierende der Medizin* [A physics laboratory for medical students]. Berlin, Germany: Logos.

Tiemann, R., Fischer, H. E., Labusch, S., & Draxler, D. (in press). Strukturierung von Lehr/Lernprozessen im Physikunterricht [Structuring teaching/learning processes in physics lessons]. In A. Pitton (Ed.), *Außerschulisches Lernen in Physik und Chemie. Gesellschaft für Didaktik der Chemie und Physik* (Vol. 23). Münster, Germany: Lit.

Tillmann, K.-J. (Ed.). (1989). *Was ist eine gute Schule?* [What is a good school?] Hamburg, Germany: Bergmann & Helbig.

Toothacker, W. S. (1983). A critical look at introductory labwork instruction. *American Journal of Physics*, *51*, 516–520.

van Ackeren, I., & Klemm, K. (2000). TIMSS, PISA, LAU, MARKUS und so weiter: Ein aktueller Überblick über Typen und Varianten von Schulleistungsstudien [TIMSS, PISA, LAU, MARKUS and so on: An overview of types and variations of large scale assessments]. *Pädagogik*, *12*, 10–15.

van den Berg, E., & Giddings, G. (1992). *Laboratory practical work: An alternative view of laboratory teaching*. Perth, Australia: Curtin University of Technology.

van Jooling, W. R., & de Jong, T. (1997). An extended dual search space model of scientific discovery learning. *Instructional Science*, *25*, 307–346.

Wagner, B. (1999). *Lernen aus der Sicht der Lernenden: eine Untersuchung zum Einfluss des Basismodell-Unterrichts auf das Lernen von Schülerinnen und Schülern; Europäische Hochschulschriften: Reihe 11, Pädagogik, Bd. 780* [Learning from the learners view: An investigation of the influence of basis model teaching on the learning of students]. Frankfurt, Germany: Peter Lang.

Waring, M. (1997). *Social pressures and curriculum innovation: A study of the Nuffield Foundation Science Teaching Project*. London: Methuen.

Weber, Th. (2002). *Kumulatives Lernen im Physikunterricht: Eine vergleichende Untersuchung in Unterrichtsgängen zur geometrischen Optik* [Cumulative learning in physics lessons. A comparative study of lessons on geometric optics]. Dissertation an der Mathematisch-Naturwissenschaftlichen Fakultät der Humboldt-Universität zu Berlin, Germany.

Weidenmann, B. (1993). Multicodierung und Multimodalität im Lernprozess [Multicoding and multimodality of learning processes]. In L. J. Issing & P. Klimsa (Eds.), *Information und Lernen mit Multimedia* (pp. 65–84). Weinheim, Germany: Psychologie Verlags Union.

Weinert, F. E. (1996). *Psychologie des Lernens und der Instruktion* [Psychology of learning and instruction]. Göttingen, Germany: Hogrefe.

Weinert, F. E. (Hrsg.). (2001). Leistungsmessungen in Schulen [Assessment of performance in schools]. Weinheim, Germany: Beltz.

Wirth, J. (2003). *Selbstregulation von Lernprozessen* [Self-regulation of learning processes]. Münster, Germany: Waxmann (im Druck).

Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist*, *24*, 345–376.

Woolnough, B. E. (1983). Exercises, investigations and experiences. *Physics Education*, *18*, 60–63.

Zoller, U. (1990). The IEE: An STS approach. *Journal of College Science Teaching*, *19*, 289–291.