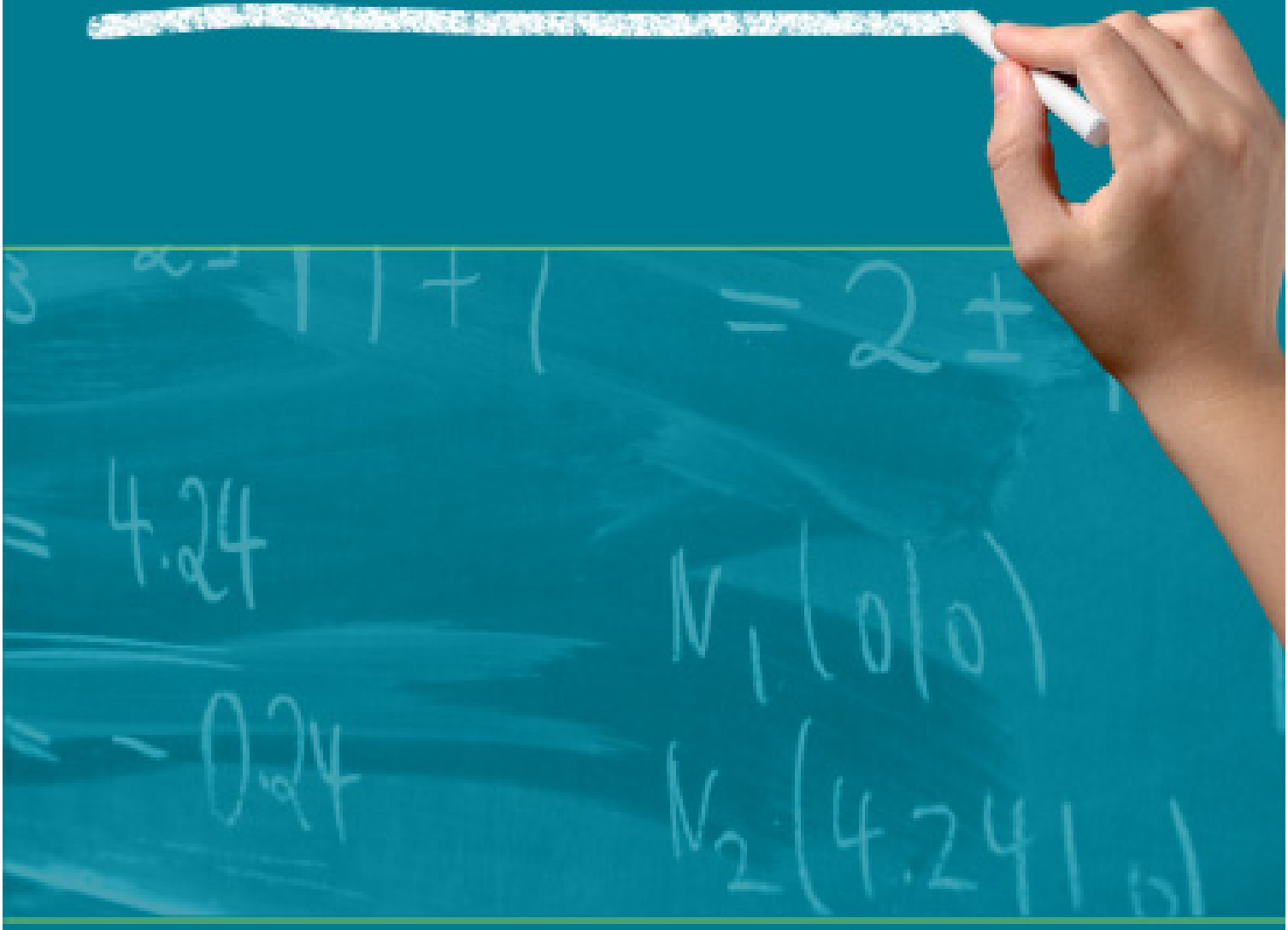
Johannes Hartig · Eckhard Klieme · Detlev Leutner (Editors)

Assessment of Competencies in Educational Contexts



HOGREFE

Assessment of Competencies in Educational Contexts

Assessment of Competencies in Educational Contexts

Johannes Hartig
Eckhard Klieme
Detlev Leutner
(Editors)

HOGREFE

Library of Congress Cataloging in Publication

is available via the Library of Congress Marc Database under the
LC Control Number 2006933006

Library and Archives Canada Cataloguing in Publication

Assessment of competencies in educational contexts / Johannes
Hartig, Eckhard Klieme, Detlev Leutner, editors.

Includes bibliographical references.
ISBN 978-0-88937-297-9

1. Competency based educational tests. 2. Educational evaluation.
I. Klieme, Eckhard, 1954- II. Leutner, Detlev III. Hartig, Johannes, 1970-

LC1034.A88 2007 371.26 C2006-904938-6

PUBLISHING OFFICES
USA: Hogrefe & Huber Publishers, 875 Massachusetts Avenue, 7th Floor, Cambridge, MA 02139
Phone (866) 823-4726, Fax (617) 354-6875; E-mail info@hogrefe.com
EUROPE: Hogrefe & Huber Publishers, Rohnsweg 25, 37085 Göttingen, Germany
Phone +49 551 49609-0, Fax +49 551 49609-88, E-mail hh@hogrefe.com

SALES & DISTRIBUTION
USA: Hogrefe & Huber Publishers, Customer Services Department,
30 Amberwood Parkway, Ashland, OH 44805
Phone (800) 228-3749, Fax (419) 281-6883, E-mail custserv@hogrefe.com
EUROPE: Hogrefe & Huber Publishers, Rohnsweg 25, 37085 Göttingen, Germany
Phone +49 551 49609-0, Fax +49 551 49609-88, E-mail hh@hogrefe.com

OTHER OFFICES
CANADA: Hogrefe & Huber Publishers, 1543 Bayview Avenue, Toronto, Ontario M4G 3B5
SWITZERLAND: Hogrefe & Huber Publishers, Länggass-Strasse 76, CH-3000 Bern 9

Hogrefe & Huber Publishers
Incorporated and registered in the State of Washington, USA, and in Göttingen, Lower Saxony, Germany

Printed and bound in the USA
ISBN 978-0-88937-297-9

Preface

The goals of education in modern societies can no longer be described by a fixed set of specialized knowledge and skills that is transferred from one generation to the next. Nowadays, knowledge must be applicable to different, new, and complex situations and contexts. It is against this background that the concept of *competence* has attracted increased attention in educational research. Competencies can be conceptualized as complex ability constructs that are closely related to performance in real-life situations. The theoretical modeling of competencies, their assessment, and the usage of assessment results in practice present new challenges for psychological and educational research. The present book covers current theoretical, psychometric, and practical issues related to the assessment of competencies in different educational settings. It is addressed to educational researchers as well as to educational practitioners and other readers interested in theoretical and practical aspects of educational assessment and evaluation, such as policy makers, teachers, and school administrators.

The work on this book took place in a broader context of political and scientific efforts to strengthen the educational research on competencies in Germany. As a result of these efforts, a priority research program on competence models and measurement founded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) was successfully established recently. Additionally, a new service structure for technology-based assessment (TBA) funded by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF) was set up at the German Institute of International Educational Research (Deutsches Institut für Internationale Pädagogische Forschung, DIPF). The TBA structure provides support for research projects using internet and computer technology as tools for competence assessments.

Parallel to the preparation of these initiatives, researchers from different countries were asked to contribute their expertise from different research fields. These contributions are compiled into the present book. They cover issues related to defining and theoretically modeling competence constructs, to psychometric measurement models, to the construction of measurement instruments (with a focus on the use of internet and computer technology), and the assessment of competencies in large-scale studies to monitor educational quality.

The first three chapters of the book provide *theoretical perspectives* on the concept of competencies in educational contexts, as well as on *developmental models*. Chapter 1 deals with the usage of the term competence and provides a working definition for competencies in educational assessments. Chapter 2 reviews the development of cognitive preconditions of successful learning and identifies turning points

Johannes Hartig, Eckhard Klieme, Detlev Leutner: Assessment of Competencies in Educational Contexts, Hogrefe Publishing GmbH, Göttingen 2008

in the development of competencies in childhood and adolescence. Chapter 3 uses the case of deductive reasoning to illustrate a model for qualitative cognitive development and the translation of this model into an adequate measurement model.

A second group of contributions deals with issues in the *psychometric modeling* of competencies, including the construction and reporting of scores, the estimation of effects of item and person characteristics, and the measurement of change. Chapter 4 describes general demands on psychometric models for the measurement of competence constructs, and illustrates different suitable psychometric models. Chapter 5 provides an introduction to the concept of explanatory item-response models, which can be used to model effects of person and item characteristics on test performance. Chapter 6 deals with linking methods to compare competence scores measured with different test forms and to measure change in competencies assessed at different points in time. Chapter 7 presents multivariate latent variable models suitable to measure multiple skills required in different combinations for complex tasks.

In the third part of the book, practical issues related to *assessment methods and technology* and to the feedback of test results are covered, with a focus on the usage of computer and internet technologies. Chapter 8 introduces fundamental quality criteria of standardized testing and describes specific demands of measurement instruments for competence constructs; Chapter 9 provides an introduction into the assessment of competencies with the help of computer and network technologies. As one of the most frequent application of computer-based testing, computer-adaptive testing is described in Chapter 10. Chapter 11 provides an overview about further possibilities of computer-based assessments of competencies with respect to innovative test and item formats. In Chapter 12, the use of internet-based assessments to support distance-learning programs is described.

Finally, several contributions deal with *large-scale assessments of competencies for the monitoring of educational quality.* Chapter 13 introduces general criteria to characterize different designs of large-scale assessments and research questions and provides an illustration based on the OECD Programme for International Student Assessment (PISA). Chapter 14 describes central characteristics of the educational standards recently developed in German-speaking countries. Chapter 15 deals with the problem of providing fair comparisons of competencies assessed in schools or classrooms differing in important background characteristics. In Chapter 16, finally, the use of internet technology to provide feedback to schools about the competencies of their students is illustrated based on a German program of comparative tests in elementary schools.

We are grateful to all who supported the production of this book, first of all authors and the BMBF, which funded the preparation of this book. Due to the preparation of the priority research program, the production of the book took longer than planned, and we would like to thank the authors, the BMBF (particularly Dorothee Buchhaas-Birkholz and Detlef Fickermann), and Hogrefe & Huber Publishers for their patience. The book covers a wide variety of questions relevant to the growing area of competence assessment in education. We hope that the contributions will

prove useful, instructive, and interesting for researchers as well as practitioners in the field.

Frankfurt am Main and Essen, May 2008

Johannes Hartig Eckhard Klieme Detlev Leutner

Contents

Johannes Hartig, Eckhard Klieme, Detlev Leutner: Assessment of Competencies in Educational Contexts, Hogrefe Publishing GmbH, Göttingen 2008

Assessment Methods and Technology

Large-Scale Assessment for the Monitoring of Educational Quality

Contributors

Eva L. Baker
National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California, Los Angeles, California, USA.
William L. Bewley
National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California, Los Angeles, California, USA.
Claus H. Carstensen
Department of Educational Science and Research Methodology, Leibniz-Institute for Science Education (IPN), Kiel, Germany.
Gregory K. W. K. Chung
National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California, Los Angeles, California, USA.
Paul De Boeck
Department of Psychology, Katholieke Universiteit Leuven, Belgium.
Lou DiBello
Learning Sciences Research Institute, University of Illinois at Chicago, Illinois, USA.
Theo J. H. M. Eggen
Cito, Psychometric Research Centre, Arnhem, Netherlands / Department of Research Methodology, Measurement, and Data Analysis, University of Twente, Enschede, Netherlands.
Ulrike Enders
Institute of Psychology, Friedrich Schiller University, Jena, Germany.
Judith Glück
Institute of Psychology, Alpen-Adria University Klagenfurt, Austria.
Johannes Hartig
Center for Educational Quality and Evaluation, German Institute for International Educational Research (DIPF), Frankfurt, Germany.
Marcus Hasselhorn
Center for Education and Development, German Institute for International Educational Research (DIPF), Frankfurt, Germany.
Ingmar Hosenfeld
Faculty of Psychology, University of Koblenz-Landau, Campus Landau, Germany.
Nina Jude
Center for Educational Quality and Evaluation, German Institute for International Educational Research (DIPF), Frankfurt, Germany.

Johannes Hartig, Eckhard Klieme, Detlev Leutner: Assessment of Competencies in Educational Contexts, Hogrefe Publishing GmbH, Göttingen 2008

Astrid Jurecka
Center for Educational Quality and Evaluation, German Institute for International Educational Research (DIPF), Frankfurt, Germany.
Eckhard Klieme
Center for Educational Quality and Evaluation, German Institute for International Educational Research (DIPF), Frankfurt, Germany.
Ulf Kröhne
Institute of Psychology, Friedrich Schiller University, Jena, Germany.
Detlev Leutner
Department of Instructional Psychology, School of Education, University of Duisburg-Essen, Essen, Germany.
Katharina Maag Merki
Institute of Educational Science, University of Education, Freiburg, Germany.
Christof Nachtigall
Institute of Psychology, Friedrich Schiller University, Jena, Germany.
Harold F. O'Neil
Rossier School of Education, University of Southern California / National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California, Los Angeles, California, USA.
Manfred Prenzel
Department of Educational Science, Leibniz-Institute for Science Education (IPN), Kiel, Germany.
Dominique Rauch
Center for Educational Quality and Evaluation, German Institute for International Educational Research (DIPF), Frankfurt, Germany.
Tina Seidel
Institute of Educational Science, Department of Educational Psychology, Friedrich Schiller University, Jena, Germany.
Christiane Spiel
Faculty of Psychology, University of Vienna, Austria.
Rolf Steyer
Institute of Psychology, Friedrich Schiller University, Jena, Germany.
Alina A. von Davier
Educational Testing Service, Princeton, New Jersey, USA.
Matthias von Davier
Educational Testing Service, Princeton, New Jersey, USA.
Mark Wilson
Graduate School of Education, University of California, Berkeley, California, USA.
Joachim Wirth
Department of Research on Learning and Instruction, Ruhr-Universität Bochum, Germany.
Kentaro Yamamoto
Educational Testing Service, Princeton, New Jersey, USA.

Part I
Theoretical Perspectives and Developmental Models

Chapter 1
The Concept of Competence in Educational Contexts[1]

Eckhard Klieme, Johannes Hartig, and Dominique Rauch

Within the social and human sciences, one area undergoing rapid expansion and attracting increased public attention is empirical educational research (Mandl & Kopp, 2005). The increasing knowledge requirements in many areas of work and life and the globalization of labor and educational markets have made the question of the educational system's productivity a crucial one for society. Since the end of the 1980s, the introduction of new oversight strategies for governmental intervention worldwide has led to a stronger focus on "outputs" and "outcomes" at all levels of the educational system, from elementary through secondary and tertiary education up to vocational and adult education. These outcomes – or the value added to them – are used as criteria for the productivity of entire educational systems, the quality of individual educational institutions, and the learning achievements of individuals. The role of educational research, then, is to render this educational productivity *measurable*, to develop models that can explain how educational process take place, evaluate their effectiveness and efficiency, and propose and analyze strategies for intervention.

In a modern industrial society, education and professional qualifications can no longer be described according to a rigid canon of knowledge in specific subjects passed on from generation to generation. Instead building competencies has been identified as the main objective of education. And while educational goals themselves are changing, the traditional methods of pedagogical and psychological evaluation – such as the criterion-oriented achievement assessments of the 1970s, which essentially translated hierarchically structured goals specific to the particular subject matter at hand into test items – are reaching their limits (Segers, Dochy, & Cascallar, 2003).

An important theoretical and practical contribution of recent educational research is the reconceptualization and operationalization of educational objectives in conceptual terms of *competence,* as well as related concepts such as literacy and life skills.

1 Sections of this chapter previously appeared in Koeppen, K., Hartig, J., Klieme, E., & Leutner, D. (2008). Current issues in competence modelling and assessment. *Zeitschrift für Psychologie / Journal of Psychology, 216,* 61–73.

J. Hartig, E. Klieme, & D. Leutner (Eds.),
Assessment of Competencies in Educational Contexts, 3–22.

The concept of competence is central to empirical studies dealing with the development of human resources and the productivity of education. Although it has been in use for decades, the term competence has enjoyed increasing currency in educational research, psychology and neighboring disciplines in the last few years (e.g., Csapó, 2004; Klieme, Funke, Leutner, Reimann, & Wirth, 2001; Klieme & Hartig, 2007, Rychen & Salganik, 2001, 2003; Sternberg & Grigorenko, 2003; Weinert, 2001).

In the first section of this chapter, three well known yet fundamentally different concepts of competence will be presented and discussed with respect to their potential to guide the measurement of educational outcomes. We will then present a working definition of competence for educational assessment settings, before we discuss current challenges in the assessment of competencies. Here we distinguish four key areas: the development of theoretical models of competence, the construction of psychometric models, the construction of measurement and research on the use of diagnostic information.

Different Concepts of Competence

In the following we will deal with the generic approach founded by Noam Chomsky, the normative approach which is prominent among educationalists and the pragmatic approaches developed by David McClelland and other scholars in psychology.

Theoretical Concepts for Linguistic Development and Socialization

In his linguistic theory, Chomsky distanced himself from the behavioristic linguistics predominant at his time, which equated language to observable sound and sentence patterns. He argued that ”If we are ever to understand how language is used or acquired, then we must abstract for separate and independent study a cognitive system, a system of knowledge and belief, that develops in early childhood and that interacts with many other factors to determine the kinds of behavior that we observe; to introduce a technical term, we must isolate and study the system of *linguistic competence* that underlies behavior but that is not realized in any direct or simple way in behavior“ (Chomsky, 1968, p. 4). Here, Chomsky proves to be a precursor to modern cognitivism, while he himself recurs to the theories of thought by Descartes and Wilhelm von Humboldt. He attributes to them the idea of conceiving language as a system of rules that enables human beings to be creative: it enables them to express ever new thoughts in ever new situations. The citation renders clear that the term *competence* is introduced as a technical term (without reference to its etymological origins) in order to describe the cognitive system underlying these creative linguistic abilities.

The question of measuring *individual* competencies actually bears no significance to Chomsky. Rather, he is concerned with understanding the cognitive basis of language-related actions that is common to all human beings. In this sense, interindividual differences only relate to *performance* as the actual realization of a competence, which is influenced by personal and situational factors and is irrelevant to the theory.

The broader sociological discussion and Habermas (1981) in particular significantly generalized the competence concept sensu Chomsky to communicative competence as the epitome of socio-cognitive rules and structures that allow individuals to generate communicative situations. This constituted a highly influential framework social scientists' use of the term *competence* up until the 1990s. For instance, Schneewind and Pekrun (1994) state in their overview of theories of educational psychology and socialization that:

> In the end, the development of communicative competencies emerges as the final goal of socialization. It intends to enable the individual to take part in the discourse of "ideal speech situations" while on the other hand it is meant to contribute towards expounding the problems associated with the social conditions that hinder an ideal speech situation. (p. 21, translation by the authors)

Generative models such as Chomsky's or similarly the Piagetian tradition (Siegler & Alibali, 2005) maintain the distinction between *competence* and *performance*. In these theories the question whether competence can be modeled and measured is identical to the question in how far it is possible to understand, describe and evaluate the functionalities of a cognitive system that *generates* contingent behavior (performance) while not being identical to it. As a rule, generative models of competence and its development are not grounded in quantitative measurements but they are determined in reconstructive manner in a broad sense. From the empirical perspective, such models are less grounded in measurements based on larger samples, but in case studies that are qualitatively reconstructed, be it hermeneutically or – as in the case of Chomsky – by formal methods.

The issue of modeling and of empirically assessing competencies has particularly been under discussion in research on language tests. Here, test concepts that address various aspects of language use repeatedly stand in conflict with Chomsky's concept of competence: "There is a difference between competence and performance, where competence equals ability equals trait, while performance refers to the actual execution of tasks" (Shohamy, 1996, p. 148; see also Beck & Klieme, 2007). By equating competencies to traits, this view opens a path to measuring individual degrees of competencies in larger groups of persons, and builds a link to the pragmatic concept of competence described below.

Competence and the Normative Concept of *Bildung*

When scholars of educational science speak about the general goals of training within modern societies, they quarrel with finding a balance between *Bildung* in the tradition of German philosophy, i.e. developing personality and allowing individuals to participate in human culture, and qualification, i.e. establishing knowledge and skills that are relevant for vocational practice. The German Heinrich Roth seems to have been the first scholar who deliberately used the notion of competence to find a compromise between the two directions. Interestingly, the introduction of the competence concept in the second volume of Roth's *Pedagogical Anthropology*, which was published in 1971, goes along with a transition from a traditional to an emancipatory concept of education. Roth defines the central objective of education as fostering *Mündigkeit* (maturity), defined as competence for responsible action. He further defines maturity as the mental constitution of a human being where heteronomy has been substituted by autonomy to the highest possible degree (Roth, 1971). Thus, he immediately connects with the enlightened tradition of education.

Roth does not give any definition for competence. Nevertheless we can assume that he was acquainted with the variants of the term in social science, and built up on them, more so because he refers to literature on the development of competence motivation elsewhere by referring to White (1959). In any case, as a psychologically trained educational scientist, Roth (1971) views competencies as individual abilities in terms of dispositions for action and judgment:

> In our view, maturity [*Mündigkeit*] should be interpreted as competence in a threefold sense: a) as self-competence – the ability to be responsible for your own action, b) professional competence – the ability to act and judge in a particular profession, and hold responsible, c) social competence - the ability to act and judge, and hold responsible, in professional or social areas that are relevant in social, societal or political terms. (p. 180, translation by the authors)

Roth's competence concept is very broad when compared with the discussion in social science. When mentioning abilities, he does not only mean cognitive dispositions for achievement, but a comprehensive ability to act that includes the affective-motivational area (Roth, 1971). Finally, the emancipatory intention associates competence with a demand for responsibility. Thus, Roth's concept of competence refers to ideal, complex goals of education, resembling a broad normative concept of *Bildung*. As a scholar who worked in empirical educational research, Roth intended to actually construct measures for this concept, but neither he nor his successors have been able to provide such measures.

A Functional-Pragmatic Concept of Competence

The functional concept of competence, used in the early seventies' psychology, is explicitly not interested in the generative, cognitive system that is independent from situations or in normative goals of education such as fostering autonomy, but instead it is interested in a person's ability to cope with challenges in particular situations. In 1973, David McClelland demanded "testing for competence rather than for 'intelligence'" (p. 1) criticizing the traditional intelligence diagnostic. While intelligence tests are deliberately de-contextualized, McClelland claimed that educational and psychological research needs concepts and assessment procedures that take into account the situation and contextualization of human action. *Competence-oriented* diagnostics was associated with a hope of improving the adjustment of test contents to real-life situations (e.g., in vocational settings), and thus being better able to predict differences in achievement in these situations. Competence according to McClelland refers to the attributes required for successfully performing particular actions. However, he does not further specify the concept with regard to any particular theory. From his perspective, any kind of individual attribute may be perceived as "competence" as far as it serves to predict success in concrete achievement: "some of these competencies may be rather traditional cognitive ones involving reading, writing, and calculating skills. Others should involve what traditionally have been personality variables, although they might better be considered competencies" (McClelland, 1973, p. 10).

Hence, the history of the subject reveals that a key feature concerning the competence concept is its stronger relation to "real life". Bandura (1990), a social psychologist, summarizes that "there is a marked difference between possessing knowledge and skills, and being able to use them well under diverse circumstances, many of which contain ambiguous, unpredictable, stressful elements" (p. 315). Connell, Sheridan, and Gardner (2003) are particularly concise in describing competence as "realized abilities" (p. 142). While intelligence research assesses cognitive achievement constructs that are generalized across a broad scope of situations, competence constructs adhere to specific areas of demands. Thus, the question: "competent for (doing) what?" is essential to any competence definition.

Nevertheless, descriptions of specific competence constructs will differ as to degrees in how far the postulated competencies can be applied across different situations. Weinert (2001) refers to *key competencies* that are characterized by a particularly broad scope of transfer, e.g., language competencies, and *metacompetencies* that facilitate the acquisition and use of specific competencies. Meta-competencies include strategies of thinking, learning, planning and governing as well as knowledge *about* tasks and strategies and knowledge of your personal strengths and weaknesses.

If competencies are regarded as context-dependent ability constructs, their development can only be conceived as resulting from learning processes where the individual interacts with his or her environment. This means competencies can be

Johannes Hartig, Eckhard Klieme, Detlev Leutner: Assessment of Competencies in Educational Contexts, Hogrefe Publishing GmbH, Göttingen 2008

acquired by learning or they even *have to* be acquired through learning, while basic cognitive abilities, in contrast, can only be learned and trained to a far lower degree (Weinert, 2001). Competencies can be acquired through experience gained from relevant situations of demand and they might be influenced by training or other external interventions, by years of practice they might be enhanced to an expertise in the respective domains. In this sense, Mayer (2003) summarizes more recent approaches of a "psychology of abilities, competencies, and expertise" as follows: "Ability can be defined as one's potential for learning knowledge that supports cognitive performance. (…) Competency can be defined as the specialized knowledge one has acquired that support cognitive performance, and expertise is a very high level of competency" (p. 265). It is thus possible to render the fact that competencies can be learned a defining characteristic of competencies against other dispositional constructs (e.g., Hartig & Klieme, 2006). Simonton (2003) characterizes competence as "any acquired skill or knowledge that constitutes an essential component for performance or achievement in a given domain" (p. 230). He illustrates this by the example of a composer who needs competencies in dealing with melody, rhythm, orchestration, dramatizing and so on.

In summary, the pragmatic psychological tradition conceives of competencies as context-specific dispositions for achievement that can be acquired through learning. Furthermore, they functionally relate to situations and demands in specific domains. The scope of these domains or of the relevant situations can vary from highly specific competencies in narrow domains to broadly conceptualized key competencies, but contextualization and learning are fundamental to all of these concepts. As will be outlined in the following section, this pragmatic, contextualized concept of competence is a useful foundation for the empirical assessment of educational outcomes.

A Working Definition of Competence in Educational Assessment

When OECD policy makers reached out to define an international program to assess the outcome of schooling, the guiding question they had in mind was what young adults at the end of education would need in terms of skills to be able to play a constructive role as citizens in society (Trier & Peschar, 1995). Thus, they crossed the boundaries of school curricula as well as the limitations of classical models of human abilities. They neither restricted educational assessment to knowledge and skills within a few school subjects nor referred to psychological theories of general cognitive abilities. Instead, they took a functional view, asking whether young adults are prepared to cope with the demands and challenges of their future life. This type of disposition for mastering unforeseen demands and tasks has been called *life skills* (Binkley, Sternberg, Jones, & Nohara, 1999) or *cross curricular competencies* (OECD, 1997; Trier & Peschar, 1995).

In fact, the functional understanding of competencies became central to the whole Program for international student assessment (PISA) as it has been implemented by

the OECD since 1998. For example, the PISA framework defines *mathematical literacy* as "an individual's ability, in dealing with the world, to identify, to understand, to engage in and to make well-founded judgments about the role that mathematics plays, as needed for that individual's current and future life as a constructive, concerned, and reflective citizen" (OECD, 1999, p. 41). Likewise, *reading literacy* and *science literacy* are related to everyday applications and authentic tasks.

As a reaction to the need for a functional approach of competencies which developed from the goals of PISA, Weinert (1999) suggested a concept of competence that should be used in large-scale assessments of educational outcome. Competencies should be defined by the range of situations and tasks which have to be mastered, and assessment might be done by confronting the student with a sample of such (eventually simulated) situations. This kind of assessment should be of greater practical use because it goes beyond compartmentalized and inert knowledge.

Among the scholars who have been working in the field of educational assessment, Richard Shavelson seems to be closest to this conception of competence, although he has not used this term systematically. Shavelson was a prominent proponent of performance assessment, i.e. assessment that incorporates hands-on, "authentic" activities based on sampling from a "population" of relevant situations and activities (Shavelson, Baxter, & Pine, 1991, 1992). He has applied this concept in diverse vocational and professional domains. Recently, Shavelson (2007) argued for a broad understanding of educational outcomes that goes well beyond academic skills and abilities measured by standardized tests:

> These additional outcomes include learning to know, understand, and reason in an academic discipline. They also include personal, civic, moral, social, and intercultural knowledge and actions – outcomes the Educational Testing Service has described as "soft". This set of outcomes – which (…) I will call personal and social responsibility (PSR) skills – are every bit as demanding as the academic skills that often get labeled exclusively as the cognitive skills and are too important not to be measured. (p. 1)

A similar understanding of "cognitive", including cognitive and metacognitive aspects of self regulation, social behavior and moral reasoning, was held by Klieme and Leutner (2006) when they – with reference to Weinert (1999, 2001) – proposed a working definition of competencies as context-specific *cognitive* dispositions that are acquired by learning and needed to successfully cope with certain situations or tasks in specific domains. This definition is strongly in line with the pragmatic definition characterized in the previous section. Competence constructs conceptualized as learnable, contextualized, cognitive dispositions are adequate criteria for the productivity of educational processes and systems. The proposed working definition of competencies not merely refers to abilities required for success within school, but claims to cover dispositions that are required for success in future, e.g. vocational situations.

Current Challenges in the Assessment of Competencies

We identify four key areas in the assessment of cognitive competencies: first and foremost, the development of theoretical models of competence (Area 1, see Hasselhorn, 2008 and Spiel & Glück, 2008, Chapters 2 and 3 in this book), complemented by the construction of psychometric models (Area 2, see Hartig, 2008; von Davier, Carstensen, & von Davier 2008 and von Davier, DiBello, & Yamamoto, 2008; Wilson, De Boeck, & Carstensen, 2008, Chapters 4 to 7 in this book). This leads onto the construction of measurement instruments for the empirical assessment of competencies (Area 3, see Eggen, 2008; Jurecka, 2008; Leutner, Hartig, & Jude, 2008; Wirth, 2008, Chapters 8 to 11 in this book). Research on the use of diagnostic information (Area 4, Chung, O'Neil, Bewley, & Baker, 2008; Hosenfeld, 2008; Klieme & Maag Merki, 2008; Nachtigall, Kröhne, Enders, & Steyer, 2008; Seidel & Prenzel, 2008, Chapters 12 to 16 in this book) rounds off the research field. In the following, we explicate the concrete questions and problems addressed within each of the four areas and outline the current state of research.

Area 1: Development of Cognitive Models of Competencies

As mentioned above, the shift towards the competence construct has prompted efforts to improve the assessment of these complex and contextualized constructs. The first question to arise here is which models provide a basis for developing measurement instruments and interpreting their results. In current educational research, only a limited number of competence models exist. Therefore, it is important to develop cognitive models that explain interindividual differences in domain-specific performance.

A first challenge in model development is the contextualized character of competencies, which means that both person- and situation-specific factors have to be taken into account. For example, when describing foreign language skills with reference to situational demands, the competencies required to read a text can be distinguished from those required to engage in conversation (e.g., by distinguishing written vs. spoken text, or text comprehension vs. text production). On the individual's side, knowledge structures relevant to different situations must be taken into account; for example, the available vocabulary, grammatical knowledge, and mastery of sociopragmatic rules (Chen, 2004; Kobayashi, 2002). This simultaneous consideration of individual- and situation-specific components has consequences for the structure of competencies as well as for the description of competence levels. Hence, two groups of theoretical models devised to describe and explain competencies can be distinguished: models of competence levels and models of competence structures (Hartig & Klieme, 2006; Klieme, Maag Merki, & Hartig, 2007). Models of competence *levels* define the specific situational demands that can be mastered by indi-

viduals with certain levels or profiles of competencies; levels of competencies are used to provide a criterion-referenced interpretation of measurement results. These models – also called "construct maps" (Wilson, 2008) – are particularly useful for assessing and evaluating educational outcomes on an aggregated level. Models of competence *structures* deal with the relations between performances in different contexts and seek to identify common underlying dimensions. These models are especially interesting for explaining performance in specific domains in terms of underlying basic abilities, and can provide a basis for more differentiated measurement results of individual-centered assessments. The two kinds of models relate to different aspects of competence constructs. They are not mutually exclusive, but ideally complementary.

The aspect of development is also very relevant in the context of theoretical competence models. To date, only a few competence models have addressed the issue of competence development (primarily in the domain of science; e.g., Bybee, 1997; Prenzel et al., 2004, 2005, Wilson, 2008). For the most part, these models still have limited empirical foundation in terms of longitudinal data, and their conceptualizations of competence development differ. Some models see competence development as a continuous progression, shifting successively from the lowest to the highest competence level (e.g., Prenzel et al., 2004, 2005). The level of elaboration and systematization increases with the competence level (as described by Bybee, 1997, for scientific literacy). Other models conceptualize competence development as a non-continuous process characterized by qualitative leaps (e.g., conceptual change in science; Schnotz & Preuß, 1997; Schnotz, Vosniadou, & Carretero, 1999). This process involves a fundamental reorganization of concepts and structures from everyday life to correspond with new science-based ideas (e.g., Vosniadou, Ioannides, Dimitrakopoulou, & Papademetriou, 2001; Wilson, 2008).

In addition, the design of cognitive models of competencies depends on the questions addressed or the decisions to be informed. A model fitting for some purposes (e.g., giving immediate feedback) may be totally ineffective for other purposes (e.g., comparative evaluation of educational institutions). A more detailed model of competencies is needed in the first case than in the second. In one case, precise estimates might be required on an individual level, in another case on an aggregated level. Switching between two purposes can cause a whole host of problems, as recent experiences in the United States have shown (Cheng, Wanatabe, & Curtis, 2004; Fuhrmann & Elmore, 2004).

To summarize, in many domains where the need for well-founded competence assessments is evident, basic research concerning theoretically as well as empirically sound models of competence structures, competence levels, and competence development is still required. Although attempts have been made to interconnect cognitive competence models with psychometric models and measurement instruments, they have often failed to meet the demands of the current, more complex definition of competencies. There is a clear need for more integrative, interdisciplinary research activities.

Area 2: Psychometric Models

As Embretson (1983) put it, psychometric models are about "modeling the encounter of a person with an item" (p. 184). Psychometric models are the link between theoretical constructs and the results of empirical assessments; they provide the measurement rules by which test scores are assigned based on performance in test situations. Given the contextualized nature and complexity of competence constructs, psychometric models for their measurement have to meet certain requirements (Hartig, 2008, Chapter 4 in this book; Hartig & Klieme, 2006). On the one hand, they have to incorporate all relevant characteristics of the individuals whose competencies are to be evaluated. Because competencies refer to performance in complex domains, the models should take into account that multiple abilities may be required. At the same time, they have to take into account domain-specific situational demands. Because competencies are conceptualized as context-specific constructs, the results of competence assessments should be related to the mastery of specific, domain-relevant situations. Item response theory (IRT) has a long tradition in educational assessment, and many of its past and recent developments were made to cater for specific needs in this area. IRT allows ability estimates and item difficulties to be compared (Embretson, 2006), thus providing a basis for models incorporating individual and situational characteristics. Several recent developments in IRT hold considerable promise for the modeling of competencies, namely explanatory IRT models, multidimensional IRT models (e.g., Hartig & Höhler, 2008), and models for cognitive diagnosis.

Explanatory IRT models (Wilson & De Boeck, 2004; Wilson, De Boeck, & Carstensen, 2008, Chapter 5 in this book) incorporate predictors for successful interactions of a person with an item, i.e. attributes of the person or features of the item ("person predictors" or "item predictors"). Specific item features can be used to represent certain situational demands. Incorporating effects of item features into the psychometric model is a highly suitable way of constructing a psychometric model of competence that takes the corresponding demands into account. Although models including item features have been in use for some time (e.g., the linear-logistic test model, LLTM; Fischer, 1997), recent developments such as the inclusion of random effects on the items side (e.g., Janssen, Schepers, & Peres, 2004; Janssen, Tuerlinckx, Meulders, & De Boeck, 2000) make them more flexible to model empirical data from complex performance situations.

Models with item features that allow situational demands to be incorporated (e.g., the LLTM) are typically unidimensional. To model performance in complex situations, it may be necessary to include more than one ability dimension in the model. A straightforward way of doing so is to apply multidimensional IRT (MIRT) models (e.g., McDonald, 2000; Reckase, 1997). MIRT models with multiple correlated scales, were each items draws on a single ability (so called "between-item multidimensionality") have been applied to data from educational assessments to take into account relations between performance in different domains (e.g., in the PISA