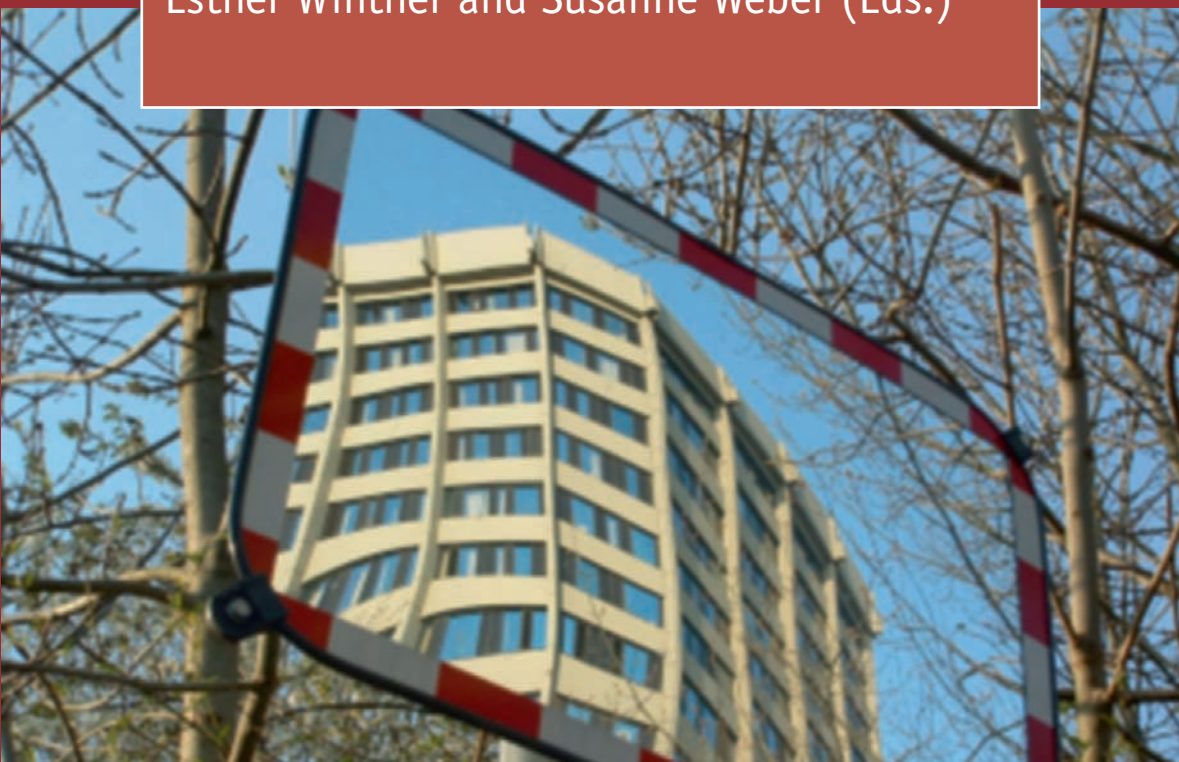


VET Boost: Towards a Theory of Professional Competencies

Essays in Honor of Frank Achtenhagen

Fritz K. Oser, Ursula Renold, Ernst G. John,
Esther Winther and Susanne Weber (Eds.)



SensePublishers

VET BOOST: TOWARDS A THEORY OF PROFESSIONAL COMPETENCIES

VET Boost: Towards a Theory of Professional Competencies

Essays in Honor of Frank Achtenhagen

Edited by

Fritz Oser
University of Fribourg, Switzerland

Ursula Renold
*Swiss Office for Professional Education and Technology,
Berne, Switzerland*

Ernst G. John
University of Göttingen, Germany

Esther Winther
University of Göttingen, Germany

and

Susanne Weber
University of Munich, Germany



SENSE PUBLISHERS
ROTTERDAM / TAIPEI

A C.I.P. record for this book is available from the Library of Congress.

ISBN 978-90-8790-736-5 (paperback)

ISBN 978-90-8790-737-2 (hardback)

ISBN 978-80-8790-738-9 (e-book)

Published by: Sense Publishers,
P.O. Box 21858, 3001 AW Rotterdam, The Netherlands
<http://www.sensepublishers.com>

Printed on acid-free paper

All rights reserved © 2009 Sense Publishers

No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

TABLE OF CONTENTS

<i>Acknowledgements</i>	ix
<i>Fritz Oser, Ursula Renold, Ernst G. John, Esther Winther & Susanne Weber</i>	
Introduction.....	xi

Part 1: Measurement on VET under an International Perspective

<i>(1) Martin Baethge, Lena Arends & Esther Winther:</i> International Large-Scale Assessment on Vocational and Occupational Education and Training	3
<i>(2) Jürgen Baumert:</i> International Comparisons: Strengths and Weaknesses and How To Overcome the Weaknesses	25
<i>(3) Richard J. Shavelson:</i> Reflections on the Assessment of Job Performance and Competence	39
<i>(4) David C. Berliner:</i> Do High-stakes Tests of Academic Achievement Limit Economic Development? The Beginning of a Long-term Natural Longitudinal Study	55
<i>(5) John Stevenson:</i> Reconciling the ‘Sacred’ and the ‘Profane’ in Developing and Assessing Competence	71

Part 2: Swiss Research on VET: The Concept of Leading Houses

<i>(6) Ursula Renold:</i> Developing Sustainable VET/PET Research in Switzerland – An Approach Combining Research, Policymaking and Practice.....	87
<i>(7) Fritz Oser & Sarah Heinzer:</i> When the Food Goes to the Horses	97
<i>(8) Uschi Backes-Gellner & Johannes Mure:</i> The Swiss Leading House on Economics of Education, Firm Behaviour and Training Policies	109
<i>(9) Jean-Marc Falter, Yves Flückiger & José Ramirez:</i> Economics of Education – Transitions, Skills and Labour Market	117
<i>(10) Albert Düggele, Sarah Heinzer, Fritz Oser & Gian-Paolo Curcio:</i> (In-)Visible Differences: On the Modelling and Validation of Competence Profiles	133

TABLE OF CONTENTS

(11) *Astrid Elke, Sandra Grieder, Corinne Tiaden, Gerhard Steiner, Christoph Metzger, Charlotte Nüesch, Fredi Büchel, Jean-Louis Berger & Nadine Kipfer:*
Fostering Learning Competencies in Vocational Education149

Part 3: Research on Learning, Curriculum, and Technology

(12) *Walter Doyle:*
Competency as a Blurred Category in Curriculum Theory175

(13) *James W. Pellegrino, Susan R. Goldman, Kimberly Lawless & Matthew Brown:*
Making Contemporary Knowledge Shareable and Useable in Teacher Education: Technology and Media Tools to Transform Teaching and Learning189

(14) *Esther Winther:*
The Effectiveness of Learning Environments or About the Difficulties to Compare Apples and Oranges.....201

(15) *Jeroen J. G. van Merriënboer:*
The Systematic Development of Competence-Based Curricula in Vocational Education and Training.....219

(16) *Detlef Sembill & Kristina Dreyer:*
Granting Time – Restricting Time – Wasting Time? Subtle Patterns in Vocational Schooling233

(17) *Josephine P. van Meer, Gerard. J. Veldhuis, Martijn L. van Emmerik & Nicolet C. M. Theunissen:*
Assistant for Instructional Design (AID): Development and Validation245

Part 4: Social and Emotional Learning: Sharing Competencies

(18) *Andreas Krapp:*
Emotional Experience and Learning Motivation261

(19) *Erno Lehtinen & Tuire Palonen:*
Social Networks and Career Development: A Case Study in a Consulting Company277

(20) *Susanne Weber, Karin Heinrichs, Verena Mayer, Stephanie Starke & Sandra Trost:*
Fit² – Fit for Business, Fit for Life: A Problem-Based Learning Module293

(21) *Åsa Mäkitalo, Christer Carlsson & Roger Säljö:*
The Art and Skill of Engraving: Craftsmanship and the Evolution of Knowing and Technologies315

(22) *James W. Pellegrino, Susan R. Goldman & Kimberly Lawless:*
Teacher Education and Technology: A Look at Current Practice and Why There is Still Much Left To Do335

Part 5: Political and Institutional Aspects of VET Research*(23) Friedrich Butler:*

To Learn Makes People Grow Large and Strong – Developing Trends in Gainful Occupation, Learning Competence and Social Inclusion in an International Perspective361

(24) Marloes Reenalda, Wim. J. Nijhof, Rolinda J. Veldkamp-De Jong & Bernard P. Veldkamp:

Economics of Education – Transition, Skills and Labour Market.....371

(25) Lorenz Lassnigg:

Policy Learning and Outcome Orientation in Austrian VET – Theoretical and Empirical Explorations385

(26) Loek F.M. Nieuwenhuis & Alan Brown:

In Search for Flexible Expertise as a Source for Firm Viability409

(27) Per-Olof Thång:

Learning in Working Life – Work Integrated Learning as a Way to Develop Competence at Work423

List of Authors437

Frank Achtenhagen: Short Curriculum Vitae443

Index.....445

ACKNOWLEDGEMENTS

This book presents refereed and revised papers of an International Conference, held in September 2007 at University of Göttingen, one of the German Excellence Universities. The Section Vocational Education and Training (*Berufs- und Wirtschaftspädagogik*) of the German Educational Research Association planned this Conference in order to honor Frank Achtenhagen who went over to the status of Emeritus after 36 years as Full Professor and Director in Göttingen. Achtenhagen is one of the leading thinkers, researchers and pathway ploughmen in the field of VET and Professional Education on one hand, but also learning and instruction in general on the other. The Conference assembled colleagues and friends from many countries who contributed by their chapters to central aspects of his current research. Some papers refer explicitly to the personal friendship of the authors and the new Emeritus.

The themes of the different papers are related to five major areas of research. A *first* topic is the work on a large-scale assessment in vocational and occupational education and training. Reason was the work of Frank Achtenhagen and his colleagues on the preparation of a VET-PISA which started in 2004 – supported by the German Federal Government – and becomes now more and more a concrete program. The contributions to this part of the book contain a project description and profound presentations and discussions of measurement and evaluation problems. It reflects also the work of Achtenhagen with respect to item response theory, measurement and testing. The *second* part of this book presents a unique endeavour of promoting VET research: The Swiss Federal Office for Professional Education and Technology (OPET) initiated a program of Leading Houses of VET which are dedicated to different important research topics. An overview is given by the Director General of OPET followed by presentations of the research strands. This program might serve as example for comparable approaches in other countries. Achtenhagen as a member of the research steering committee of this comprehensive program is part of its foundation. The *third* part highlights central issues of research on learning processes, curriculum theory and the use of technology – also focussing on consequences for teacher and trainer education. Here the work of Achtenhagen on competence-profile modelling, competence measurement and instructional and curriculum designing is touched. The *fourth* part stresses social and emotional learning as important aspects of VET learning – together with the focus on its cognitive dimension. The *fifth* part considers the political and institutional dimensions of VET research which have to be taken into account when an international large-scale assessment of VET shall be getting started. Achtenhagen's work on learning at work, leaning in economics and learning under the conditions of institutional modelling are discussed.

The whole book is a collection of central issues around a field that is not yet taken enough into consideration in educational research, but pushed and supported by Frank Achtenhagen: VET research. He belongs to the founding fathers of it, and this is why precisely the book reflects new trends and new concepts with respect to the question “What can we, educational psychologists and educational researchers, learn from a rich and central research field.”

ACKNOWLEDGEMENTS

The Editors would very much like to thank all authors for their excellent contributions and hope that this book may enrich a goal-centered discussion of VET problems. They thank the German Research Foundation, the German State of Lower Saxony and the University of Göttingen for their support of the Conference and OPET for its support of this book.

Fribourg, Berne, Göttingen, Munich October 2008

Fritz Oser, Ursula Renold, Ernst G. John, Esther Winther & Susanne Weber

FRITZ OSER, URSULA RENOLD,
ERNST G. JOHN, ESTHER WINTHER & SUSANNE WEBER

INTRODUCTION

The contributions to this book are placed under one heading: research on VET competencies under an international perspective. It is effected with the overarching goal to prepare a VET-PISA, a large-scale assessment, for selected European countries. The different chapters focus on important aspects of such an endeavour. All authors are highly acknowledged experts out of different disciplines whose expertise covers the respective thematic range. The essays are structured under five headings.

The first part treats explicitly problems of measurement of VET under an international perspective. The idea of a “VET-PISA” was initiated in 2004 by a call of tender of the German Federal Ministry of Economy and Labor. Frank Achtenhagen and Martin Baethge won the tender and formed a working team of specialists. This success was mainly founded on their interdisciplinary cooperation: Achtenhagen specialized on the micro level of the VET system, whereas Baethge overtook the work on the meso and macro levels. The team was completed by Lena Arends, Edvin Babic, Volker Baethge-Kinsky, Susanne Weber and Esther Winther and published in 2006 a feasibility study for the German Federal Government (see Baethge et al., 2006). In the following, the German Federal Ministry of Education and Research became sensitive for this project. - Further chapters of this first part investigate important details of how to proceed as the conditions in the fields of VET are very different compared to those of the PISA and TIMSS studies for general education. Martin Baethge, Lena Arends and Esther Winther present major aspects of the actual work on this topic. Jürgen Baumert as one of the leading experts for running the TIMSS and PISA studies in Germany comments on the project from the background of his rich scientific and political experiences. Richard Shavelson does the same – exploiting his broad and deep expertise of research on large-scale assessment in schools, jobs and the military. David Berliner is warning the audience with regard to possible negative consequences out of large-scale studies which become quasi-naturally independent and cause very often the contrary to the intended educational main effects. John Stevenson demonstrates how theoretically driven construction of competence concepts might help overcoming such negative consequences – taking into account the different possibilities of expressing the “meaning” of measures and behaviour. All chapters of this first part present and discuss decisive aspects – including philosophical and methodological issues – of large-scale assessment studies, under special focus on VET.

The second part presents a unique field experiment for VET: the Swiss system of Leading Houses. It is the trial to establish sustainably research structures and to develop also careers of young researchers. Overarching goal is to establish a stable organization for VET research. Fritz Oser has been the President of the escorting

Steering Committee and responsible for the structure of this convincing concept. Frank Achtenhagen followed him in 2008. Ursula Renold as Director General of the Swiss Federal Office for Professional Education and Technology which is heading the Leading House system presents the central goals of this approach linking science, policy and VET practice. Fritz Oser and Sarah Heinzer visualize the research criteria for the Leading House concept – followed by presentations of the different Leading Houses. Uschi Backes-Gellner and Johannes Mure report on the Leading-House on Economics of Education, Firm Behaviour and Training Policies, focussing on a more micro-economic perspective whereas the other Leading House on Economics of Education, managed by Jean-Marc Falter, Yves Flückiger and José Ramirez, stresses mainly macro-economic problems of VET and labour markets. Albert Düggele, Sarah Heinzer, Fritz Oser and Gian-Paolo Curcio present the work of the Leading House “Quality of Vocational Education and Training” which is especially dedicated to the education and training of VET teachers and VET trainers. A very innovative accent is set by the construction of teaching vignettes which serve as stimuli for the training of teachers and trainers. Members of three working groups under the roof of one Leading House cast their eye on problems of VET learning: Astrid Elke, Sandra Brieder, Corinne Tieden and Gerhard Steiner; Christoph Metzger and Charlotte Nietsch; Fredi Büchel, Jean-Louis Berger and Nadine Kipfer. In total, this Swiss approach might serve as example for a concentrated, fruitful promotion of VET research also for other countries which spend much and even more money – absolutely and relatively – on VET projects without this stringent policy.

A third part of this book is dealing with research topics on learning opportunities, curriculum, instructional design and technology. The construction, implementation and evaluation of complex learning environments are central for the work of Frank Achtenhagen and his team. Parallel to the endeavour on the promotion of situated learning their virtual enterprise becomes part of the official State curriculum and serves as basis for about one year of instruction in the subjects Business Administration and Accounting in German Commercial Schools. The following contributions stress important aspects and open questions of the international discussion. Walter Doyle demonstrates why it is necessary to sharpen the understanding of the construct “competencies” with regard to the development of an efficient curriculum theory. James W. Pellegrino, Susan R. Goldman, Kimberly Lawless and Matthew Brown discuss basic problems of the reform of teacher education programs by introducing technology. Esther Winther shows results of a field study by which special effects of computer-based learning environments have been measured with regard to the influence of coaching guidance on the one side and in comparison with traditional classroom settings on the other. Jeroen van Merriënboer presents new approaches to foster VET individualized by media-based modes of selecting authentic tasks to be fulfilled and to find, thereby, own learning trajectories. It is also shown how the learning results are sampled by protocol portfolios. Detlef Sembill and Kristina Dreyer are concentrating on the problem of time which has eminent importance for the organization but also the learning opportunities within VET. Problems of how much time should be spent in enterprises or how much time is provided for

vocational and academic subjects define the characteristics of each VET system. Josephine van Meer, Gerard Veldhuis, Martijn van Emmerik and Nicolet Theunissen discuss instructional design with regard to teaching support for teaching personnel without a special educational training within the Dutch military. Major problem to be solved is job rotation within military which urges a fast offer of knowledge needed for the new job.

The fourth part is treating aspects of social and emotional learning under the special focus of sharing competencies. Frank Achtenhagen always tries to also take into account motivational and emotional variables into his field research. Thus, he also fostered corresponding dissertation and habilitation projects. Andreas Krapp discusses basic assumptions of how the theoretical concepts of emotion and learning motivation shall be related to each other for getting better hypotheses to be proved empirically. Erno Lehtinen and Tuire Palonen describe how individual development and social interaction during the acquisition and maintenance of expertise in an ill-defined field can be integrated. A model of social network analysis serves as evaluation instrument of this process. Susanne Weber, Karin Heinrichs, Verena Mayer, Stephanie Starke and Sandra Trost report on the results of a problem-based learning program for promoting key-competencies (as e.g. argumentation or teamwork/collaboration) within a BA program of Business Administration. They demonstrate how basic assumptions of this approach are validated. Åsa Mäkitalo, Christer Carlsson and Roger Säljö present a case study on the VET conditions and the practice of the craftsmanship of engraving under the central aspect of evolution of knowledge and technologies. Their focus on the aesthetic and artistic expression within this occupation enriches the VET discussion which is actually run mainly under globalization and modernization aspects. James W. Pellegrino, Susan R. Goldman and Kimberly Lawless discuss in their second contribution to this book possibilities to strengthen modes of teacher education by introducing intelligently technology.

The fifth and last part is dedicated to political and institutional aspects framing VET research. Frank Achtenhagen worked and is still working as political advisor in the field of VET for many national and international institutions. The contributions start with Friedrich Buttlar's statement on necessary support for learning at all levels as precondition for effective and efficient VET – based on his experiences as President of a large research institute (IAB), as Vice-Minister and as responsible manager of ILO. Marloes Reenalda, Wim Nijhof, Rolinda Veldkamp-De Jong and Bernard Veldkamp discuss effects of VET learning in Dutch higher professional colleges in dependency of dual learning environments. Central question is how the characteristics of the different environments influence the development of competencies. A study for Austria on how policy itself can learn for the fields of education and training and how such learning processes are concentrated on selected VET problems is presented by Lorenz Lassnigg. Major finding is that learning results in a special field do not necessarily lead to a clear political decision. Loek Nieuwenhuis and Alan Brown treat the topic of flexible expertise of the workforce by studies on four labour organisations (The Netherlands) and small ICT and engineering companies (United Kingdom). This concept is seen as necessary condition for continuous innovative processes and as

basis for targeted investment in human resources under the auspices of lifelong learning. Per-Olof Thång discusses the problem that on-the-job-learning is not enough any longer for competence development. People at the worksite need also theoretical reasoning and reflection, thus, focussing on the interaction and interconnection of case-related abilities and systematic embedding of case-related knowledge, thereby, urging new ways to teach and train for occupations.

All chapters of this book show important facets for the development of VET and VET research. All the chapters are also connected to the outstanding and fundamental work of Frank Achtenhagen. Many of them mirror his thinking in a substantial way. The editors are sure that this book will help to develop further these topics.

PART 1:
MEASUREMENT ON VET UNDER AN
INTERNATIONAL PERSPECTIVE

INTERNATIONAL LARGE-SCALE ASSESSMENT ON VOCATIONAL AND OCCUPATIONAL EDUCATION AND TRAINING¹

1 Vocational and Occupational Education and Training (VET) in a Globalized World

The economy but also the social and political conditions of life are minted by so-called “megatrends” (cf. Buttler, 1992; Achtenhagen, Nijhof, & Raffe, 1995; Achtenhagen, & Grubb, 2001). We have to cope with an increasing internationalization and globalization of economic exchange relations on the markets for goods, services and labor, the European nations´ growing together, the increasing importance of knowledge within the work processes, the progressive use of information and communication technologies. These processes urge an adaptation of the present workforce to the changes of work structure. As the effects of the “megatrends” hit every national economy, the states have to find out adequate and efficient ways to deal with them, especially and mainly

- how to organize the transition from school to work,
- how to reform and adapt the given system of vocational and occupational education and training (or even to install such system),
- how to organize modes of lifelong learning.

In principle, we find five different types of how to cope with problems of VET, mainly how to treat young people who are not on the direct “royal road” to Higher Education: (a) One tries to collect these students in “vocational” schools which mainly adopt the character of “schools for the residuals”. (b) One tries to convey as much academic education as possible. The hope in this case is that only few adaptation processes to any job will be necessary. A typical example is the differentiation of Bachelor training (in the UK: “fast track degrees”) – which for Germany e.g. would compete with an apprenticeship of articulated clerks or comparable apprenticeships. (c) One tries to offer as many positions as possible for semi-skilled jobs or simple work activities. (d) One tries to install modes of partnership between enterprises, schools and government for developing models of VET which are to be seen as first steps of introducing an apprenticeship model. (e) Few countries – Austria, Denmark, Germany, Luxembourg, Switzerland – try to stay with their dual system of apprenticeship but have to be vigilant as the “megatrends” also cause trouble for this well developed type of VET. One example is the supply of posts for apprenticeships with its dependency on the current status

¹ The idea of a VET-PISA was initiated by a call of tender of the German Federal Ministry of Economy and Labour. The feasibility study was finished in 2006 (Baethge et al.). The following activities are supported by the German Federal Ministry of Education and Research. Frank Achtenhagen was and is centrally involved in all corresponding activities with the plan to start formally in 2009.

of economy: E.g., the number for Germany is at the moment at 43% of the corresponding cohort; 17% visit vocational full-time schools and about 40% enter so-called “measures” (or “transition courses”) by which the government tries to provide some vocational preparation and simultaneously to improve academic compulsory education. The participants of these “measures” – a lot of them are migrants – show very low grades and also severe language problems which hinder them to get an apprenticeship (cf. Baethge, Solga, & Wieck, 2007). One severe consequence of the decline of the number of apprenticeships is a lack of highly qualified workers and employees – which on the side of the government and the economy strengthens the endeavour to increase the number of posts and to generate also reduced modes of apprenticeship with regard to prior knowledge and language proficiency for getting a majority of the youth into the labor system and corresponding qualification processes.

These different types of dealing with VET problems are also challenged by the need for new qualifications, i.e. to design new formats of professional profiles and new combinations of competencies. However, whereas the agents in economy act by and large autonomously the VET systems are – depending on different national regulations – more or less governed by states and communities which react to the new developments more or less quickly, more or less comprehensively and more or less competently. Thus, the megatrends initiate different, opposite reactions: One leads to harmonization, adaptation and adjustment of VET systems to economics worldwide. The second goes into the direction of new differences between VET systems due to the different national political decisions on how to react to the new challenges which overarch the whole world.

Under a labor market but also an individual perspective these processes are accompanied by an old problem which has become dramatically virulent in quality and quantity: unemployment. At least in the Western European states the labor force needed for industrial production and, partially, also for traditional services has been reduced to a large extent because of different reasons, not the least in order to boost shareholder value. It is not only an economic question of how to deal with the millions of people to be supported financially by the state. A serious problem is also the mental health of these people, their acceptance by those who are employed, by their families and peers. It is a very new upcoming task of VET to prepare especially young people for phases of, if not for a life in unemployment and, nevertheless, to qualify them for societal necessary work.

Along with this a general task of VET and also of economy is coming to the fore. From the point of view of general education the most important objective is to arrive at all the new qualification targets and economic goals not only by making people useful for the work process but also by helping them to become mature, authentic individuals conducting sovereignly their lives, as well as to develop them to attentive, critical and constructive citizens, actively participating in social and political processes. This is an objective of education in itself, a humanistic task which lies not only in the interest of both, individuals and society, but also in the interest of the economic subsystem which cannot prosper in a social environment where people merely function as producers and consumers.

2 Necessities and Possibilities of an International Comparison of VET

The European Commission has seen these severe problems. With regard to economic and social conditions of life, it has put forward the ambitious economic and social goal of becoming “the most competitive and dynamic knowledge-based economy in the world” (European Commission, 2005). Therefore, improved participation in education and the labour market is playing a crucial role in reaching these goals.

The practical implications of the problems of competence measurement are reflected in the ongoing progress towards the development of a standardized European Qualifications Framework (EQF). In Europe, the ministers for vocational education and training in 32 countries have accepted the Maastricht Communiqué, endorsed in December 2004, consisting of an agreement for the development of an overall European Qualifications Framework (EQF) and a European Credit Transfer System for VET (ECVET). The aim of the EQF is to provide a common set of reference levels as an integration of ECVET and the European Credit Transfer System (ECTS) in Higher Education (Commission of the European Communities, 2005). The main objective of the EQF is to provide a meta-framework for the development of a set of common descriptions for qualifications that can be applied in all educational systems in Europe. All countries involved try now to translate the EQF into a NQF – a National Qualifications Framework – which shall function on the national as well the international level. The comparison of qualifications necessarily leads to the focus on the teaching/training and learning/working dimensions of VET with regard to the development of occupational competencies. This is the decisive dimension of VET compared to the institutional/organizational dimensions. Under a political and economic perspective as well as under an educational one it is important to see how the different VET systems help young people to acquire the necessary competencies mentioned above. That is the main task with regard to the fulfilment of the overarching objectives of the EQF and the NQFs.

Therefore, a central task is to assess those competencies on the micro-level in an objective, reliable, and valid way, as well as to link the results to the macro- and meso-structural factors (the following remarks fall back upon ideas expressed by a feasibility study for a PISA-VET: Baethge et al., 2006). These factors, in turn, impact the value of competencies at a later point in time in the labour market.

Such an assessment of VET should be done as large-scale assessment for different countries in a comparable way – with regard to policy-pragmatic as well as scientific aspects:

From a *policy perspective*, one can expect an improvement of knowledge related to steering and controlling of VET, particularly with regard to

- the relation between individual/biographical characteristics, educational programs, and competence development;
- the link between the outcomes of comparative competence measurement and institutional orders of VET systems (revealing strengths and weaknesses of different VET organisations in different countries);

- the interrelation between certificates of final examinations and the competencies assessed; and
- the classification of different examinations in VET in international systems of classification (ISCED, EQF) to improve acknowledgement methods of certificates at the European level.

From a *scientific point of view* one can expect gaining deeper insights into generating hypotheses as well as a valid and reliable mix of methods for an internationally comparable longitudinal study of different VET organisations in a large-scale assessment. The development of a valid and reliable frame of methods to compare the performance of different national VET systems could be considered an independent scientific goal, in particular of a cross-sectional pilot study. Moreover, from a content-related point of view, one can expect to gain deeper insights into the relationship between

- the level of competencies, educational programs, and context variables;
- different competence dimensions (generic, cross-occupational and vocational subject-specific); and
- vocational and occupational competencies and outcomes in the labour market and employment.

An international comparison of vocational education and training must be based on a common understanding of the goals of VET. This common understanding cannot be implicitly postulated, but must be mutually developed from a scientific research and policy point of view. On the one hand goals for VET can be based on a relatively narrow approach, which is focused on the required workplace skills. It has to be mentioned that this restricted view does not cover modern approaches of education neither the economy of training the workforce. Therefore, on the other hand, we need a broader approach, which incorporates, in addition to subject-related competencies, those skills individuals need to participate effectively as members of a flexible, adaptable, and competitive workshop and in lifelong learning. In accordance with the ongoing scientific discussion, there are three central goals, which educational systems must address at the system level:

- The development of potentials for individual's occupational mobility, self-regulation and autonomy;
- the safeguarding of human resources in a society, and
- the warranty of social participation and equal opportunities.

These goals function as reference points for the definition of competencies, which must be developed in vocational education and training. The first goal, *individual vocational adjustment*, denotes the ability of individuals to develop relationships with their environment and to create their educational pathways and life in society in a responsible and self-directed way. Individual vocational adjustment also denotes generic competencies like self-management skills, problem-solving skills, communication skills, and meta-cognitive skills. Individuals are considered within the context of individual aims and efforts on the one hand, and beneficial and obstructive environmental conditions on the other hand. This perception includes the goal of VET systems to provide opportunities for developing and constructing individual and occupational identities.

The second goal, *safeguarding of human resources*, subsumes every aspect of educational systems that facilitates individual abilities to act at work and in the

labour market (individual economic user perspective) and helps to provide workforce requirements (social demand perspective). It refers to the suitability of vocational education and training and the development of occupational systems (e.g., Achtenhagen, & Grubb, 2001; Baethge et al., 2003). Moreover, we can distinguish between rather quantitative and qualitative aspects: From a quantitative point of view VET systems ought to supply occupational systems as best as possible – i.e., avoiding narrow professional qualifications or over-qualifications in little-demanded or seldom available domains. From a qualitative point of view, an adequate preparation for labour market requirements, which includes subject-related competencies as well as generic skills, is the main focus. The scientific discussion labels both as specialisation versus generalisation, which is also a focus in the context of lifelong learning and continuing education.

The third goal, *warranty of social participation and equal opportunities*, emphasizes the relationship between VET and social structures. This refers to the contribution of VET to minimize dependencies between social background and educational, life, and income opportunities, and to enhance social integration and participation of young people in processes of shaping the social and political community.

To measure and compare the performance of national VET systems regarding the three goals and to be able to relate them to institutional structures and input criteria, there are at least three methodological problems:

- First, how can vocational competencies be measured and compared?
- Second, how can the relevant micro and macro-structural conditions of VET systems in different countries be analysed and compared?
- Third, with regard to the differences of job classification schemas in the participating countries, how can occupational fields and work activities be identified?

These three aspects shall be shortly discussed.

3 The Competence Dimension of VET

One of the main research questions of an international comparison of VET is to determine how suitable measurement tools of competence can be identified.

Competence measurement in the field of VET is more complex than in compulsory education. Whereas international large-scale assessments like TIMSS (Third International Mathematics and Science Study) and PISA (Program for International Student Assessment) are limited to assessing mathematics and science performance of fourth- and eight-graders (TIMSS) or literacy, numeracy, science, and problem-solving performance of 15-year olds (PISA), a comparable VET study has to take into account individuals' performance in the workplace and in the labour market as well as practical aspects (e.g. motor skills, dexterity). Moreover, international student assessment programs like TIMSS and PISA are based upon well-grounded research traditions and internationally validated concepts, like a world curriculum for mathematics. By contrast, a VET study cannot draw on comparable concepts concerning the structure and development of vocational expertise in various occupational fields. The difference between measurement in

VET and compulsory/academic education can be described by focussing on a variety of aspects:

(1) Measurement in the field of VET is always action-and activity-related.

- A reliable and valid measurement of competencies has to take into consideration, in addition to the cognitive dimensions, the affective and psycho-motoric dimensions as well as the corresponding personality traits (Achtenhagen, & Grubb, 2001).
- Recent research and development shows that the link between cognitive goals and knowledge to metacognitive functions gains more and more importance (Anderson, & Krathwohl, 2001).
- Another important distinction refers to the trait- and state-perspectives of personality. “Trait” is related to relatively stable personal attitudes whereas “state” represents the situation-related aspects of behaviour (cf. Winther, 2006). However, as opposed to state attributes trait attributes do not sufficiently explain the variation of achievement or the level of competencies. Therefore, the measurement of state components should be incorporated in a VET large-scale assessment.

(2) Actions and activities in the fields of VET are very specific – not only with regard to the national scope of occupations but also regarding the differences between different countries and occupations themselves:

- The variety and heterogeneity of occupational specialisations, even within one society, makes it very difficult to reach an international agreement concerning consistent competence standards. For example, an analysis in the field of car repair at the European level revealed fundamental differences: Whereas in one country only one individual is necessary for mending an entire car, in another country more than ten different experts are responsible for fixing different parts of the car. Thus, if neglecting this difference, an international comparison in the field of car repair would have generated the wrong conclusions.
- In the context of VET, very little research has been done to develop comparable, internationally valid concepts of the structure and development of vocational expertise in different occupational fields. Additionally, there is no international agreement on competence-based occupational standards or levels. At the European level an ongoing progress towards the development of a standardized European Qualifications Framework (EQF) and a European Credit Transfer System for Vocational Education and Training (ECVET) could provide a basis for further developments in this field. A VET study could complement these approaches and provide solutions on the basis of well grounded scientific support.
- Due to differences between national labour markets and cultures, the impact of curriculum-based objectives has to be reconsidered: The majority of VET systems is not based upon formal qualifications (e.g., Germany, Austria, Switzerland) and curriculum contents function as a basis for entering the labour market. Thus, from a political and institutional point of

view, the question of how to compare curriculum contents and objectives is one of the central aspects in a VET study, taking into consideration aspects of assessment and research on learning and instruction.

- In accordance with the three broad goals for VET, a VET study must focus on at least four different competence levels and their interrelationships:
 - General subject-specific competencies, like reading, writing, mathematics, and problem-solving, have been discussed for a long time in the context of international comparative studies of compulsory education.
 - Generic occupational competencies are related to successful performance in the labour market. They refer to the notion of key skills or “core competencies”, which comprise knowledge about the structures of organizations and labour markets, interacting in socially heterogeneous groups, acting autonomously, and using tools interactively (Rychen, & Salganik, 2003).
 - Vocational subject-specific competencies refer to the collection of required individual prerequisites in order to successfully develop occupational identities and solve work-related tasks.
- domain-specific skills refer to successfully carried out occupation-specific tasks (these skills contain occupation-specific contents, job-specific as well as enterprise specific skills – see Oates, 2004).

(3) Internationally well-known schemata to describe competencies in the fields of VET focus on cognitions and the cognitive domain but do not take into consideration differentiations of dimensions of knowledge.

- International research in the area of learning psychology agrees that the knowledge dimension of learning goals and contents must be taken into consideration for an adequate description of learning processes and achievement (Bransford, Brown, & Cocking, 2000). This aspect is also supported by the revision of Bloom’s taxonomy of cognitive objectives where the formal classification of learning goals is combined with different knowledge dimensions (Anderson, & Krathwohl, 2001).
- Without an extensive description of knowledge dimensions, which are embedded in vocational and occupational actions and activities, the necessary competencies cannot be adequately recorded.
- It is recommended to differentiate declarative, procedural, and strategic knowledge (cf. Achtenhagen, 2004). Declarative knowledge corresponds to factual knowledge as knowledge of facts, structures, and knowledge nets. The domain of procedural knowledge subsumes the application of knowledge, i.e. how to operate with facts, structures, knowledge nets and their corresponding elements. The development of strategic knowledge is considered one of the central goals for VET.

(4) Vocational and occupational actions and activities require always competencies with regard to occupational orientation, awareness of workplace tasks, worker-participation, and social relations. These competencies are subsumed under the heading “employability”.

Another important question is how to relate competencies of compulsory/academic education to competencies of vocational/occupational education. This is a crucial aspect from an educational policy point of view: Is it sufficient to achieve superior qualifications in compulsory/academic education before entering the labour market or is it necessary to enrol in an apprenticeship within the dual system, which comprises 3-4 days of weekly work in an enterprise and attending 2-1 days in a vocational school (“Berufsschule”)? There is no doubt that learning on the job is necessary to support the “transition from school to work” but very little is known about the advantages and disadvantages of the different arrangements.

Due to the differences between measuring competencies in VET and compulsory/academic education it is very difficult to identify a schema for describing and measuring competencies in VET without severe overlapping of categories. Baethge, et al. (2006, p. 45) developed a schema of how to measure competencies (table 1).

Self-competence

Self-competence contains every aspect of personal development in a narrow sense – the development of cognition, emotion, motivation and ethics (Achtenhagen, 1996, p. 27). In the context of an international VET study, personal characteristics which impact vocational performance (attitudes, values, perceptions, incentives, motivation, and meta-cognitive strategies) have to be measured.

Cognitive Competence

Recent empirical studies show that a differentiation of knowledge according to declarative, procedural, and strategic knowledge is plausible and feasible (e.g., Bransford, Brown, & Cocking, 2000). In the domain of declarative knowledge (factual knowledge), new concepts, which are related to work requirements, have to be taught, trained and learned. Particularly with regard to the increasing complexity and networked structures in work contexts, declarative knowledge cannot be developed and fostered in linear, abstract, and fragmented ways.

Second, the domain of procedural knowledge subsumes the application of knowledge relating to the way of operating with knowledge facts, structures and nets and the corresponding elements.

Third, the development of strategic knowledge is considered one of the central goals for VET. Here, the focus is on behavioural conditions and problem solving, which refers to applying deep declarative and procedural knowledge, taking into account not only possible primary effects but also unintended side effects. Thus, decisions are made with full awareness of possible negative consequences. This seems to be valid for all sectors and levels of VET.

To measure aspects of strategic knowledge, complex tasks have to be presented. Individuals are expected to weigh arguments and to come to a well-founded solution, taking into consideration intended outcomes as well as possible unintended side-effects. Typical examples, i.e. for the fields of business, might be decisions on purchase, sales or production planning.

Functional Competence

To illustrate the relevant underlying occupational subject-related competencies in different educational tracks and occupational tasks, an example of empirical research, consisting of significant occupational tasks in German occupations has been discussed (Baethge, & Baethge-Kinsky, 2006). Central results for the different occupational profiles have been:

- Subject-related knowledge is fundamental in every occupational field;
- knowledge domains play a substantial role in every occupational field, with regard to processes of social interaction, analytical and reasoning skills, meta-cognitive skills, and organisational participation;
- in each occupational domain a sense of perception of typical critical situations is being developed; and
- sensory-motor skills are central part of work behaviour (in particular in the field of skilled work in industrial production and trade).

Table 1
Description of competence domains

INDIVIDUAL CAPACITIES (accessed and interpreted in different contexts)	Competence Domains (different contextualised areas of performance)		
	SELF-COMPETENCE		
	COGNITIVE COMPETENCE	FUNCTIONAL COMPETENCE	SOCIAL COMPETENCE
	(theoretical/analytical requirements) “applying concepts”	(technical/practical/ functional requirements) “using tools, equipments and technical resources”	(interpersonal requirements) “interacting with others”
ATTITUDES VALUES PERCEPTIONS	Self-efficacy for predicting achievements (4 items, applied in the PISA survey: Kunter, et al., 2002, p. 168) “Big five” (10 Items)		
INCENTIVES MOTIVATION	Meta-cognitive strategy: effort and perseverance (4 items, applied in the PISA survey; Kunter, et al., 2002, p. 171) Interest and motivation (18 items, Prenzel, et al., 2001; short version applied in a national project funded by the German Research Foundation, DFG) State components of motivation: Preparedness of effort and endurance; Mastery ambitions; Acting on task valence (12 items; Winther, 2006). Meta-cognitive strategy: self-related cognition (verbal) (3 items, applied in the PISA survey; Kunter, et al., 2002, p. 169)		

METACOGNITIVE STRATEGIES	LIST-Questionnaire (inventory of measuring learning strategies) (22 items, Wild, & Schiefele, 1994; short version applied in a national project funded by the German Research Foundation, DFG, $\alpha = 0,74-0,81$)	
DECLARATIVE KNOWLEDGE	Complex tasks for simultaneous measurement of cognitive and functional competence	Questionnaire for mindful identity negotiation
PROCEDURAL KNOWLEDGE		(17 items, Weber, 2005, $\alpha = 0,89$)
STRATEGIC KNOWLEDGE	According to the occupational field: complex task for measuring cognitive competence complex task for measuring functional competence	(consists of items for declarative, procedural, and strategic knowledge in the domain of social competence)
GENERAL ABILITIES	Literary; numeracy	

Social Competence

Social competence refers to individual preconditions necessary for behaving and interacting with others. With regard to measurement questions (especially external versus internal measurement) it will cause problems to collect data on social competence by questionnaires. Instruments to measure this competence dimension by referring to the construct of “mindful identity negotiation” might be helpful (cf. Weber, 2005).

External Versus Internal Measurement

A comparative VET study has to cope with the fact that, corresponding to the following measurement considerations, research results show that there are two basic approaches for the identification of competencies – based on external performance as opposed to internal conditions:

- Measurement of competencies by watching working behaviour (performance), or
- measurement of competencies on the basis of tests which can include work samples (internal conditions).

There are four arguments disadvising the application of the first approach for the purpose of identifying competencies in a VET study:

- Measurements of performance of specific working tasks entail an agreement of performance levels in different occupations with regard to occupational tasks. Moreover, the tasks must be considered relevant in every participating country. It is difficult to agree upon typical occupational tasks/bundles of activities across countries.
- Performance-based measurement is very time-consuming, in particular with regard to reaching a high level of validity, reliability, and objectivity. For example, the working samples applied in the World Skills

Competition comprise up to 20 hours, which is only feasible given the low number of participants in the contest.

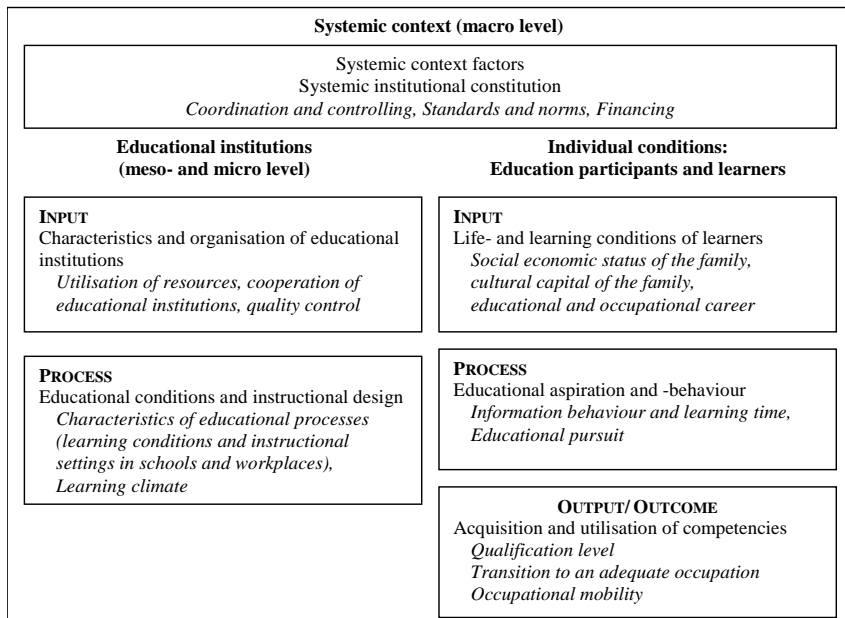
- Superior goals of VET cannot be incorporated sufficiently. Aspects of work-related qualifications are included, whereas facets of personal development and societal participation tend to be excluded.
- Work psychology also argues against the external measurement approach because measurement of performance in working processes is too complex and time-consuming (cf. Lang-von Wins, 2003).

4 Institutional and Individual Conditions for Quality in VET

With regard to the micro-, meso- and macro-levels which can be differentiated for VET systems there is a set of indicators representing significant information about the quality in the organization of VET. Tables 2 and 3 (cf. Baethge et al., 2006, pp. 70, 86) illustrate the relevant context factors in a coherent framework, which at the same time reflects the objectives and the stages of activity and achievement (input, process, output), social contexts, and educational structures (context and input factors).

Table 2

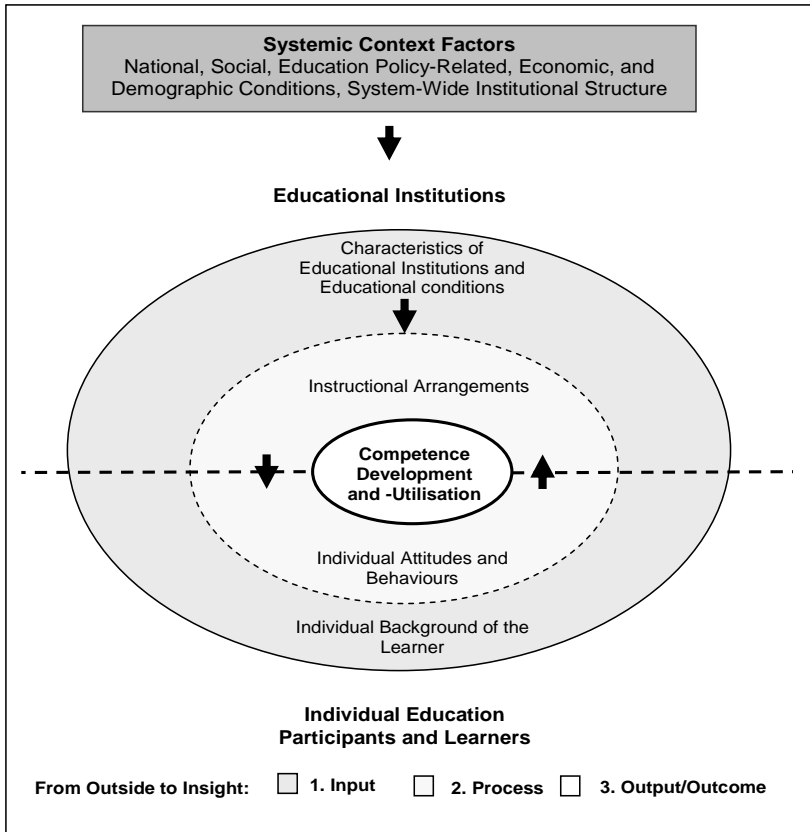
Context factors for quality in vocational education and training



The need for overall information on education- and employment-systems is based on the assumption that a comparison of different initial VET programs in the context of a VET study can only be explained against the background of the corresponding social, education policy-related, economic, and demographic

conditions of the respective country. There are different economic, social, and political context factors impacting vocational education and training and the corresponding conditions for acting: At a macro-level of the education- and employment-systems, we distinguish between systemic context factors (e.g., social, cultural, economic, and political conditions) and systemic institutional constitutions (coordinating and steering, standards and norms, and financing of VET systems). Institutional constitutions must provide the necessary conditions for steering and coordination as a basis for adjustments and arrangements of qualitative and quantitative educational offers with regard to the demand of individuals and enterprises (Descy, & Tessaring, 2002; Lassnigg, 2000). Steering instruments consist, on the one hand, of different forms for providing material resources to ensure and develop educational offers and, on the other hand, of standards and norms to condition the contents and methods of educational processes.

Table 3
Correlations between institutional and individual conditions for the development and utilisation of competencies



At the macro-level of the institutional frame (inputs) we have to look for indicators to analyse

- institutional arrangements defining standards and curricula in VET (in particular the role and involvement of different actors as governments, providers, social partners etc.);
- differences in VET curricula in terms of narrow workplace-related and broad cognitive skills;
- financing systems of VET (by state and/or enterprises);
- standards for qualification requirements of teaching staff and continuing professional development; and
- regulations for public quality control of institutions and offers of VET.

At the (meso and micro) levels of educational institutions, indicators relating to the structural and organisational conditions of schools and firms play an important role. They more or less pre-form the learning and teaching processes (Kunter, & Stanat, 2002). At the input level this refers to the characteristics and organisation of educational institutions and providers of educational services (utilisation of resources, co-operation of educational service providers, quality control). At the process level this relates to the instructional setting and the learning environment (learning conditions and instructional settings, such as self-directed learning or teamwork in schools and workplaces, learning climate).

The quality of learning in schools or firms refers to the relationship between implemented pedagogical and organisational settings (e.g., self-organised learning, project-based learning, or teamwork) and to the results of competence measurement. In addition, the type of institution (e.g., library) as well as supplementary activities provided by schools or firms (e.g., special needs education), must be identified and analysed with regard to their utilisation by students. In this regard, it can be assumed that an intensive use of such learning opportunities correlates with high learning achievements.

Important indicators for the quality of instruction could be (Onstenk, 2003; Nieuwenhuis, 2004):

- Opportunities for content-related experiences: What kind of content-related experiences, such as definitions, theories, and problems, are accessed when performing working tasks?
- Complexity of working tasks: To what level of proficiency is a working task related (e.g., aspects of complexity, pre-structuring, problem-orientation, personal responsibility)?
- Opportunities for task-related, socially interactive discussion: What types of opportunities for task-related, socially interactive discussions with employees or experts are provided?
- Problem-orientation and complexity refer to the cognitive demands of work (complexity of cognitive operations, unfamiliar tasks, number of pre-structured solution steps).
- Variation: learners/workers should be involved in different aspects of the work process in order to acquire a feeling for the big picture, which is a precondition for being able to effectively transfer knowledge.

- Participation: to enhance critical reflective work behaviour; it is important that workers/learners perceive that they can participate in workplace decisions.
- Support: scaffolding the learning process by foremen or expert colleagues is an important aspect of the learning potential for students/apprentices at the workplace.

The category “individual background” mainly refers to correlations between family background and students’ performance. Educational outcome is influenced by the family background in many different and complex ways. Thus, the socio-economic status of families can be considered an important variable for explaining variances in student achievement (Baumert, & Schümer 2001). For example, the PISA results show that the parental occupational status, as a measure of socio-economic status, can influence students’ aspirations and attitudes regarding learning.

At the input level learners’ individual background variables may be evaluated by using questionnaires. In following international comparative studies we need indicators for the

- socio-economic status of family (including migration status);
- social and cultural capital of family (see Baumert, & Schümer, 2001);
- educational career of the learner;
- participation in non-formal and informal learning activities in work and social contexts as well as self-organized learning activities (in particular with regard to little formalized VET systems; informal learning has a significant influence on the development of general and vocational competencies); and
- spare-time activities as well as the corresponding integration in social networks.

In addition to the background variables we need indicators regarding learning behaviour, aspirations and the perception of training processes:

- information behaviour (provides insight into the scope of knowledge, which is the basis of career choice during the transition from school to VET);
- learning activities (indicate learning achievement in schools which could be operationalized according to the average amount of time invested in learning activities);
- career (vocational) aspirations (individual’s projects in terms of action plans for realizing individual’s objectives);
- perception of vocational training processes experienced in VET (only by a set of indicators concerning the perception of training within the questionnaires it is possible to gain information on the quality of instructional settings and procedures).

Supplementary to measuring vocational competencies a VET study must provide information on possibilities for utilising vocational competencies developed in VET in workplaces and labour markets, and with regard to individual educational career developments. Individual outputs, however, are not solely based on VET processes and personal characteristics; they are rather determined by aspects of the respective occupational field and by organisation-specific conditions.

This refers to opportunities during the transition from VET to work, and to the assessment of vocational competencies in the labour market. Thus, the following attributes can only be evaluated on the basis of a detailed examination and documentation of systemic context factors.

For the questionnaire there is a need of indicators for labour and labour market outcomes, i.e.

- formal level of vocational qualifications (including horizontal and vertical combinations of different educational systems in terms of access authorisations, i.e. for entering college or university);
- transition to an adequate job (this indicator refers to the quality of career entry after completing VET);
- quality of the workplace achieved in terms of
 - complexity of working tasks;
 - personal responsibility;
 - scope of action;
 - participation in workplace decisions;
 - status within the firm;
 - learning opportunities at the workplace;
- occupational mobility in terms of perceived discontinuity as well as of career mobility.

The data concerning individual social background, inputs, outcomes and learning experiences have to be ascertained by interviews with the students/apprentices involved.

5 Modes of Realisation of an International Comparison of VET

Compared to school-based comparative studies (e.g., PISA, TIMSS), a comparison of VET is different regarding institutional, as well as content-related aspects. Despite the differences in micro-structural aspects of learning, compulsory education can refer to a common curriculum of basic competencies, such as literacy, numeracy, science, and comparable age groups in the classroom. In contrast, VET varies to a large extent in content, given the large diversity of occupational fields and the institutional heterogeneity within and between societies. Therefore, an international comparison of VET entails the measurement of cognitive, as well as functional, aspects of competence. The complexity of a large-scale assessment can be characterised as a two-fold problem of comparison:

- Vertical comparability refers to the educational level and can be determined by institution (e.g. tertiary level (academic track), secondary level) or by duration (number of years) and/or age.
- Horizontal comparability refers to differences in curricula and occupational fields and to the problem of defining comparable vocational tracks.

For the purpose of linking the results of an individual competence assessment to the micro- and macro-structural context factors, both aspects of comparability are essential and have to be taken into account. From a methodological point of view, a comparison of national samples at the same educational level, incorporating the

same competence requirements, involves a number of problems. Vertical comparability is problematic since it refers to national traditional backgrounds of educational systems. Horizontal comparability, on the other hand, refers to national, specific, occupational profiles and the corresponding curricula within the existing structures of labour organisation (Heikkinnen, 2001).

Vertical Comparability

From an international point of view, the identification of comparable VET programs is based on very few and little updated data, mainly provided by the following institutions: OECD, UEO, or EUROSTAT/CEDEFOP. The most beneficial sources for identifying quantitatively-relevant programs of initial VET were published by EUROSTAT and CEDEFOP a handbook of different educational programs in different European countries-based on data from 1995-1996 and updated in 2000. It consists of abstracts containing information about different educational programs, the type of program (vocational preparation, general education, vocational education and training), ISCED level, and typical entrance age. Moreover, information regarding the type of labour market qualification (generic, subject specific) and the number of participants is provided. Unfortunately, the same information is not available for every country and data on the number of beginner participants entering educational programs are entirely missing.

Based upon data analyses and discussions with international experts, we come to the conclusion that ISCED is not a very suitable classification for identifying initial VET programs addressed to a medium level of proficiency, since it relates to levels of general education, rather than VET. According to ISCED, the corresponding educational programs in different countries are located at levels 3 and 4 (with exceptions found on level 5). Provided that ISCED level 3 represents the lower limit for VET programs in the sample, certain vocational preparation and short-term programs¹ would be located below these categories.

Horizontal Comparability

The problem of horizontal comparability refers to identifying comparable occupational field and educational tracks in different countries. An initial identification could be based on the following four broad occupational fields:

- Industrial/technical occupations in industry and trade;
- Commercial and commercial/administrative occupations in commerce and other services;
- healthcare occupations in the field of personal services; and
- information and communication technology in the field of information/technical services.

¹ Germany: Basic Vocational Education (Berufsgrundbildungsjahr, einjährige Berufsfachschule); Denmark: Basic Vocational Education (EGU); Switzerland: „Anlehre“; Hungary: Secondary VET Schools; USA: Vocational High Schools; UK: NVQ-level 1 or 2.

However, there are still problems remaining: Firstly, there is neither an internationally consistent competence-based concept of the structure and classification of initial VET programs/occupational fields, nor a reporting to rely on (Lisbon-to-Copenhagen-to-Maastricht Consortium Partners, 2004). This does not imply that classification systems are completely missing, but what is needed are basic data in a usable format. At an international level alongside ISCO 88 (for occupational fields), an occupational classification of discrete fields of training has been developed on behalf of CEDEFOP (Andersson, & Olsson, 1999). However, neither of the two classifications has been accepted as part of the ongoing reporting at the European level. National differences of institutional interrelations between education and employment systems on the one hand, and the impact of non-standardised vocational biographies, on the other hand, require considering the following aspects:

Incorporating non-standardised vocational biographies is only feasible on the basis of retrospective measurement of educational experiences within a cross-sectional study. The sample should be based on employed population with criteria including age and occupational tasks according to level and breadth-to reconstruct vocational biographies. Representativeness in this approach is limited due to institutional aspects of education; for example, the “fuzziness” of retrospective measurements of level and content of different VET programs.

Provided that VET programs, levels, and institutions highly correlate with occupational fields and levels (e.g., in Germany based on the Berufsprinzip), and at the same time, cover most occupational tasks at a medium level: vertical (based on occupational tasks), or horizontal (based on VET programs), sampling does not seem to make a big difference to the outcome. However, even in Germany, this correlation does not apply in every occupational field, and in Anglo-Saxon countries there are very few segments of this type. With regard to a panel study starting from educational institutions, a sampling based on the structure of occupational tasks is recommended. Nevertheless, for pragmatic reasons, a sample based on the structure of VET programs has to be taken into consideration. The differences between the two options are due to the remaining efforts in preparing the sampling and evaluating the weight of each sample in different segments. Both options are connected to the construction of vocational sub-samples.

First Option: Broad Sample Based on Occupational Tasks

This method is based on occupational tasks to be performed at a particular age (e.g., 24 years). The sample is defined on the basis of the “European Labour Force Survey”. The ISCO 88 classification (International Standard Classification of Occupations) provides a basis for approaching a comparable sample within the most important occupational fields (industrial and manufacturing, commercial and administrative, social and healthcare, information and technology). ISCO 88 has been implemented within the “European Labour Force Survey” and indicates not only the level, but also the contents of occupational tasks. In this way, aggregates of homogeneous occupational contents and levels, which are representative (at least in Germany) of a considerable number of individuals in the same age, could be identified. (Table 4): With the exception of skilled data processing workers

(constituting a proportion of 1 percent of information technology occupations), the sample represents the most important occupational field of employees, comprising about 45 percent of young adults in this age. The information value of ISCO data can be considered a solid basis for the level and content of occupational tasks, even though occupational fields are not always explicitly distinguishable. Thus, an adjustment procedure for the purpose of a representative identification of bundles of tasks must be conducted.

The adjustment procedure is based on an analysis of work requirements according to the main functions: production, distribution, personal services, and focus of contents. The job profiles could be used as a basis in an iterative procedure for sampling in which they provide reference points for identifying national tasks corresponding to the respective profiles. This allows adequate classifications of the tasks that are related to adjoined levels or different bundles of tasks according to ISCO 88. Problems of classification help to identify clearly distinguishable occupational profiles at an international level as a basis for developing measurement tools.

Table 4
 Selected occupational tasks, ISCO codes, and employment rates in different occupational fields at the age of 15 to 25 years (current employment rate in Germany)

Segment	Occupational Field	ISCO Code of Specific Occupational Tasks	Employment Rate in Different Occupational Fields (15 to 25 Years) (in %)
Industrial/Technical Occupations (Industry and Trade)	Metal and Electrical Work	72	15,0
Commercial/Administrative Occupations	Commercial and Office Work	522, 523, 3416, 3419, 411, 343	9,7 11,9
Social, Education and Healthcare Occupations	Nursing	322, 323	8,8
Information Technology Occupation	Skilled Data Processing Workers	312	1,0
Total			100,0

The bundles of tasks could be combined with national data on educational background variables derived from national labour force surveys, to identify the (quantitatively) most relevant VET programs and levels in the corresponding occupational field. The identified VET programs are the foundation for institution-based sampling. However, this method is particularly problematic with regard to Anglo-Saxon countries since broad sampling includes a variety of institutional arrangements that have to be accounted for. A stratified sample would have to account for different institutions and regional differences, as well as for the identification of homogeneous age groups.

Second Option: Narrow Sample Based on Widespread and Well-defined VET Programs

This method is based on VET programs with comparable curriculum contents, preparing young people between 16 and 20 years to enter the labour market. This method consists of two steps: First, the development of two schemata, one consisting of a pre-selection of designated VET programs (Table 5) and another based on EUROSTAT/CEDEFOP data, supplemented by information on relevant VET programs from non-European countries. Second, based on this extract, a proposal for sampling, incorporating the main occupational fields and at the same time providing a basis for comparing level and age, should be developed.

The field of trade is represented by the European occupational profile of car-mechanics, of which a similarly-structured profile can be found in the U.S. and Australia. The field of industrial metal occupations is represented by electricians and mechanics (industry), although the latter may possibly be replaced by tool makers at some point). Contrary to Germany, vocational qualifications of bankers, skilled computer scientists, and nurses are formally located at a higher proficiency level in most other countries. However, with regard to the contents of curricula they appear comparable.

Table 5
VET program, corresponding ISCED levels, age range at entrance

Occupational Field	VET Program	ISCED Levels (Internationally)	Age Range at Entrance
Metal and Electrical Occupations	Car Mechatronic	3	16–20
	Mechanic (industry)	3	16–20
	Electrician (industry)	3	16–20
Commercial/ Administrative Occupations	Banker	3–5	17–20
Healthcare Occupations	Nursing	4–5	18–20
Information and Communication Technology Occupations	Skilled Computer Scientist	3–5	17–20

The sampling option outlined does not provide a solution for every problem. However, aspects with regard to comparing levels (e.g., nursing) could be solved relatively easy on the basis of discussions with international experts and statistic centres.

Despite all the remaining uncertainties regarding an appropriate age range for entering a VET program, there are possibilities for identifying an age-homogeneous overall sample. The relevant age would be 18 to 19 years,—a measuring point at the beginning of VET in this age cohort (e.g., 18 year old beginners). Due to the specific starting points of each method (broad occupational tasks versus single VET program curricula for comparatively-adjusted

employment), the probability of identifying additional relevant VET programs, covering the same occupational tasks, is relatively low. Therefore, the selection method of the second option needs to be checked on the basis of a task analysis of educational backgrounds.

Possible Realization of a VET Large-scale Assessment

Taking into consideration the goal of comparing competencies and their development in VET in different countries and at the same time measuring the effects (outcomes) of these competencies after entering the labour market, a longitudinal design with three panel waves is recommended:

- The first wave should take place shortly after entering a VET program to measure the initial competence level of students.
- The second wave should take place shortly before completing a VET program and after approximately 2.5 years (provided the study is of a 3-year duration) to measure competence development during VET.
- The third wave should take place three years after completing a VET program and after the transition into the labour market to evaluate whether the individual is working in a job, enrolled in a program of continuing education, or unemployed. Measurement is focused on the sustainability of competencies, in addition to their expansion and usability in the domains of work and everyday life.

Measurement of competencies will refer to generic, cross-occupational, and subject-related vocational dimensions (Table 1), taking into consideration institutional and individual contexts of learning (Table 2 and 3). Intending to involve the entire variance of vocational competencies requirements, institutional and biographical arrangements etc. into the research design would not be realistic because the entire variety of occupational fields cannot be incorporated. From a pragmatic and methodological research point of view, a stratified sample is therefore recommended. A sample construction comprising a relevant (such as important employability segments and institutional arrangements) as well as comparable extract (for example “middle” level of competencies, similar content-related focus of knowledge and skills, same age), of initial vocational education consisting of young people between compulsory education and work. Measurement will have to be conducted in quantitatively relevant occupational fields in order to assess the major tracks of competence development.

References

- Achtenhagen, F. (1996). Entwicklung und Evaluation ökonomischer Kompetenz mit Hilfe handlungsorientierter Verfahren - am Beispiel der Ausbildung zum Industriekaufmann/zur Industriekauffrau (Skizze eines Forschungsprojekts). In W. Seyd, & R. Witt (Eds.), *Situation – Handlung – Persönlichkeit. Kategorien wirtschaftspädagogischen Denkens. Festschrift für Lothar Reetz* (pp. 137-147). Hamburg: Feldhaus.
- Achtenhagen, F. (2004). Prüfung von Leistungsindikatoren für die Berufsbildung sowie zur Ausdifferenzierung beruflicher Kompetenzprofile nach Wissensarten. In M. Baethge, K.-P. Buss, & C. Lanfer (Eds.), *Expertisen zu den konzeptionellen Grundlagen für einen*

- Nationalen Bildungsbericht – Berufliche Bildung und Weiterbildung/Lebenslanges Lernen* (pp. 11-32). Bonn, Berlin: BMBF.
- Achtenhagen, F. in Cooperation with J. W. Nijhof, & D. Raffé (1995). *Feasibility Study: Research Scope for Vocational Education in the Framework of COST Social Sciences*. European Commission: Directorate-General Science, Research and Development, COST Technical Committee Social Sciences, Vol. 3. Brussels, Luxembourg: ECSC-EC-EAEC.
- Achtenhagen, F., & Grubb, W.N. (2001). Vocational and Occupational Education: Pedagogical Complexity, Institutional Diversity. In V. Richardson (Ed.), *Handbook of Research on Teaching, Fourth Edition* (pp. 604-639). Washington, DC: American Educational Research Association.
- Anderson, L. W., & Krathwohl, D. R. (with Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., et al.) (Eds.) (2001). *A Taxonomy for Learning, Teaching, and Assessing. A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.
- Andersson, R., & Olsson, A.-K. (1999). *Fields of education and training*. Luxembourg: Eurostat. http://www.minedu.sk/igrsvs/igr_doc_20020904_2_ISCED97_manual.rtf
- Baethge, M., Achtenhagen, F., Arends, L., Babic, E., Baethge-Kinsky, V., & Weber, S. (2006b). *PISA-VET – A Feasibility-Study*. Stuttgart: Steiner.
- Baethge, M., & Baethge-Kinsky, V. (2006). Ökonomie, Technik, Organisation: Zur Entwicklung von Qualifikationsstruktur und qualitativem Arbeitsvermögen. In&(Eds.). *Handbuch der Berufsbildung*. 2. Aufl. Opladen: Leske&Budrich.
- Baethge, M., Buss, K.P., & Lanfer, C. (2003). *Konzeptionelle Grundlagen für einen Nationalen Bildungsbericht – Berufliche Bildung und Weiterbildung/Lebenslanges Lernen*. Bonn: BMBF.
- Baethge, M., Solga, H., & Wieck, Markus (2007). *Berufsbildung im Umbruch – Signale eines überfälligen Aufbruchs*. Berlin. Friedrich-Ebert-Stiftung.
- Baumert, J., & Schümer, G. (2002). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzwettbewerb im internationalen Vergleich. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele et al. (Eds.), *PISA 2000: Die Länder der Bundesrepublik Deutschland im Vergleich* (pp. 159-202). Opladen: Leske + Budrich.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How People Learn: Brain, Mind, Experience, and School*. Washington, D.C.: National Academy Press.
- Buttler, F. (1992). Bildungslandschaft bis 2010. In F. Achtenhagen, & E.G. John (Eds.), *Mehrdimensionale Lehr-Lern-Arrangements* (pp. 162-182). Wiesbaden: Gabler.
- Descy, P., & Tessaring, M. (2002). *Kompetent für die Zukunft – Ausbildung und Lernen in Europa. Zweiter Bericht zur Berufsbildungsforschung in Europa: Synthesebericht*. Reihe Cedefop reference series, Bd. 5. Luxemburg.
- Europäische Kommission (2002). *Beschäftigung in Europa 2002 – Jüngste Tendenzen und Ausblick in die Zukunft*. Brüssel: http://europa.eu.int/comm/employment_social/news/2002/sep/employment_in_europe2002_de.pdf
- Heikkinen, A. (2001). *Ist eine stärkere Berücksichtigung der historischen und kulturellen Dimension in der Vergleichenden Berufsbildungsforschung erforderlich? In CEDEFOP/DIPF (Eds.), Vergleichende Berufsbildungsforschung in Europa. Ansätze, Politikbezüge und Innovationstransfer* (pp. 118-134). Thessaloniki.
- Kunter, M. et al. (2002). *PISA 2000 – Dokumentation der Erhebungsinstrumente*. Max-Planck-Institut für Bildungsforschung, Berlin. <http://www.kmk.org/schul/pisa/Datensaetze/Gesamt-Druckversion.pdf>
- Lang-von Wins, T. (2003). Die Kompetenzhaltigkeit von Methoden moderner psychologischer Diagnostik-, Personalauswahl- und Arbeitsanalyseverfahren sowie aktueller Management-Diagnostik-Ansätze. In J. Erpenbeck, & L. von Rosenstiel (Eds.), *Handbuch Kompetenzmessung* (pp. 585 – 618). Stuttgart: Schäffer-Poeschel.

- Lassnigg, L. (2000). *Steuerung, Vernetzung und Professionalisierung in der Berufsausbildung*. Wien: Institut für Höhere Studien (IHS). http://www.lebens-langes-lernen.at/downloads/IHS_SteuerungBerufsausbildung_1100.pdf
- Nieuwenhuis, L. (2004). *The learning potential of the workplace*. http://edu.fss.uu.nl/ord/full-papers/Nieuwenhuis-L_L%204_2.DOC
- Oates, T. (2004). The Role of Outcome-based National Qualifications in the Development of an Effective Vocational Education and Training System: The Case of England and Wales. *Policy Futures in Education, Volume 2 (1)*.
- Onstenk, J.H.A.M. (2003). *Work based learning in continuing vocational and professional education, Paper for WACE - First Draft*. http://www.wacerotterdam2003.nl/documents/final_papers_abstracts/140.doc
- Prenzel, M., Kramer, K., & Drechsel, B. (2001). Selbstbestimmt motiviertes und interessantes Lernen in der kaufmännischen Erstausbildung – Ergebnisse eines Forschungsprojekts. In K. Beck, & V. Krumm (Eds.), *Lehren und Lernen in der beruflichen Erstausbildung. Grundlagen einer modernen kaufmännischen Berufsqualifizierung* (pp. 37-63). Opladen: Leske + Budrich.
- Rychen, D.S., & Salganik, L.H. (2003). A holistic model of competence. In D.S. Rychen, & L.H. Salganik (Eds.), *Key Competencies for a Successful Life and a Well-Functioning Society* (pp. 41-62). Cambridge, Göttingen: Hogrefe & Huber.
- Weber, S. (2005). *Intercultural learning as identity negotiation*. Frankfurt am Main: Lang.
- Wild, K.-P., & Schiefele, U. (1994). Lernstrategien im Studium. Ergebnisse zur Faktorenstruktur und Reliabilität eines neuen Fragebogens. *Zeitschrift für Differentielle und Diagnostische Psychologie, 15*, 185-200.
- Winther, E. (2006). *Motivation in Lernprozessen. Konzepte in der Unterrichtspraxis von Wirtschaftsgymnasien*. Wiesbaden: DUV.

INTERNATIONAL COMPARISONS: STRENGTHS AND WEAKNESSES – AND HOW TO OVERCOME THE WEAKNESSES

1 Universalization and Standardization of the Modern School

One of the great human achievements of the past century was the worldwide universalization of school education at secondary level. Talcot Parsons, who was the first to diagnose this trend, spoke of a worldwide educational revolution (Parsons, 1971); Meyer, Ramirez, and colleagues tracked and documented this trend in detail (Meyer, Ramirez, Rubinson, & Boli-Bennett, 1977; Meyer, Ramirez, & Soysal, 1992; Ramirez, & Boli, 1987). Parallel to its universalization, the time structures, social organization, and contents of schooling underwent extensive standardization (McEneaney, & Meyer, 2000). Cross-national and cross-cultural comparisons of school structures and programs reveal startling similarities in the ways schools are organized, in the delineation of learning areas and subjects, in curricular programs and working conditions, in teachers' academic training, and even in their professional views. Content is most highly standardized in the less culture-dependent subjects, with mathematics being the prime example. Baker and LeTendre (2005) have provided compelling evidence for this process of standardization drawing on data from the Third International Mathematics and Science Study (TIMSS).

In a globalizing world, where knowledge and human capital are the crucial resources of the future, the universalization and standardization of schooling injects an element of competition that no country can afford to ignore. Given the scarcity of resources, moreover, the principle of accountability is being extended to education worldwide, with ever more significant effects. Competition and accountability are global trends with unforeseeable outcomes. As one element in this process, international comparisons of education systems have gained a firm place in educational research and politics (Wiseman, & Baker, 2005).

2 Strengths of the Latest Generation of International Educational Assessments

For many years, the Educational Testing Service (ETS) in the United States was the intellectual parent and mentor of international educational assessments. There can be no doubt that the development of international large-scale assessments would have taken a different and less successful path had it not been for the support of the ETS. With the latest generation of studies conducted by the

Organisation for Economic Cooperation and Development (OECD), however, international educational assessments have finally come of age in terms of both their theoretical and programmatic conception and their technical implementation.

The theoretical frameworks guiding test development in the latest generation of studies are exemplary in all domains. The reading literacy framework of the Program for International Student Assessment (PISA) draws directly on research on cognition as well as experimental research on reading comprehension (Artelt et al., 2001; Kintsch, 1998; Kirsch, Jungebluth, & Mosenthal, 1998; OECD, 1999). At the same time, it covers a broad range of reading tasks and activities, reflecting the demands faced by young people in the modern world. It is perhaps regrettable that literature and aesthetics are underrepresented in the PISA framework, but the available test items suffice to show that German educational institutions underperform in these domains as well (Artelt, & Schlagmüller, 2004).

PISA's theoretical frameworks for mathematics and science literacy are also highly commendable, representing a faithful implementation of modern standards – as specified for mathematics in the National Council of Teachers of Mathematics (2000) standards, for example, and for science in the American Association for the Advancement of Science (1998) standards. The test items developed on the basis of this framework are exemplary in many respects. Half of the items have a multiple-choice format; the rest are open questions – some requiring short answers, others calling for extended argumentation. Many of the items reflect authentic instructional practices. As envisioned by mathematics and science experts, the assessment is thus well aligned with regular classroom instruction.

The development of item response theory scaling and multimatrix models means that assessments can now cover broad curricular domains at all proficiency levels and that task difficulty and individual ability can be represented on a single metric (Fischer, & Molenaar, 1995). Item response theory has two major advantages over classical test theory that have doubtless helped to promote the broad acceptance of the latest generation of international educational assessments. Because modern test theory allows item difficulty and individual ability to be represented on a single metric, it is now possible to estimate the probability of a given test item being solved correctly at each ability level. Specific test items can thus be used to describe which operations a student of a certain ability level is likely to be able to perform – and which are likely to be too difficult. This common metric is the basis for a hermeneutic interpretation of ability levels. It is this anchoring in terms of the content of school knowledge and the high quality of test items that makes the assessments interesting for teachers, administrators, and politicians alike. The assessments are gradually beginning to tell a story about specific national strengths and weaknesses, and this story can be communicated to a broader public.

The validity of an item-response model further implies local stochastic independence of the individual items. In other words, the estimation of an individual's ability is no longer dependent on the specific selection and arrangement of items in a calibrated test. Once the item parameters are known, participants no longer have to complete the whole test but can be presented with selected tasks from the item pool. It is thus possible to administer a large number of items covering broad curricular domains without overtaxing individual participants. The assumption of local stochastic independence of items is always violated in reality, however, and

its validity is limited. Researchers therefore tend to rotate item clusters or test booklets rather than single test items in multimatrix sampling designs.

It was thanks to PISA's solid theoretical underpinning, the quality of its items, the anchoring of its results in terms of content, and the multimatrix sampling design that the PISA findings not only attracted widespread attention in Germany – a country that is traditionally hostile to these kinds of assessments – but triggered lively discussion and debate on educational reform, both on the political level and within schools.

Another methodological strength of the latest generation of international educational assessments is the quality of their sampling procedures. Compared with earlier studies run by the International Association for the Evaluation of Educational Achievement (IEA), sampling quality is now so high that findings can, in most cases, be justifiably generalized to the target population. There have been particular efforts to ensure the adequate representation of at-risk groups of weak learners and minorities. Although there is still room for improvement in specific instances, the overall quality of sampling must be commended.

A final strength of the new generation of assessments is the stability of the research structures and the quality of the international infrastructure provided by the OECD. Because participation in PISA was endorsed and coordinated at government level, country-specific commitment has remained consistently high. Compared with the opportunities afforded by the IEA's more casual approach, the OECD's primarily political orientation has proved highly effective.

3 What Has Been Achieved by Implementing These Instruments in International Student Assessments?

Three main accomplishments can be identified:

- (1) The findings provide benchmarks of what can realistically be achieved in compulsory schooling by the age of 15 years – benchmarks with which national standards can be aligned. The regular assessments implemented in the OECD program allow performance in individual countries to be monitored against international standards.
- (2) The second strength of these studies is that they provide substantive insights into the distribution of student performance. Besides performance distribution relative to the mean, another interesting feature is the anchoring of the assessments in terms of content, which allows researchers to identify at-risk populations who lack the skills and abilities needed to succeed in the world of labor and to describe these groups in terms of their social and ethnic characteristics. By investigating the size and composition of these at-risk populations, we can gain insights into an education system's performance in the most critical ability ranges. The size of this low-achieving group is particularly important from the social perspective. In this respect, the PISA results were an unwelcome surprise for a number of European countries. Moreover, PISA showed that the U.S. school system performed better, in terms of limiting the size of this at-risk group, than has often been assumed.

- (3) Finally, the OECD studies were the first comparative educational assessments to take a state-of-the-art approach to the assessment of student backgrounds (Baumert, & Schümer, 2001). The assessments of social status, cultural and social capital, and immigration background were exemplary, and the results were surprising. Western European countries that considered themselves to be prime examples of social equalization in fact showed greater social disparities in student knowledge acquisition than did either the United Kingdom or the USA, both of which tend to be seen as prototypes of social inequality among the Western states. A cross-country comparison of social gradients, an illuminating measure of the relationship between student social background and proficiency, shows that the approaches that are most effective in targeting both the enhancement of performance levels and the reduction of inequality are those focusing on the lower performance ranges. From the perspective of raising and leveling the bar, and disregarding, for the time being, the accountability system linked to it, the No Child Left Behind program clearly has the right focus.

The findings of a comparative review on the performance of immigrant students in PISA 2003 have also had a corrective effect (Stanat, & Christensen, 2006). The data reveal an alarming need for urgent action in a number of countries – and challenge some widespread beliefs. There are, for example, marked differences between Germany and the United States. Whereas the United States has been remarkably successful, by international comparison, in its efforts to integrate young people of the second generation, the German-speaking countries have failed to address the problems of immigration and now face considerable difficulties that show no signs of abating in the second generation (Walter, & Taskinen, 2007).

Of course, it is equally important to recognize the conceptual weaknesses and structural limitations of international educational assessments – only then can their shortcomings be addressed and their misuse in the political and scientific spheres be minimized, if not prevented.¹

4 Conceptual and Methodological Weaknesses of International Educational Assessments

One would have to read the international PISA reports very carefully to notice that the age cohort of 15-year-olds and the PISA target population – that is, all 15-year-olds who are still at school – are not always identical and that the discrepancy varies considerably from country to country. In PISA 2000, the age cohort and target population were identical in countries such as Japan, where all 15-year-olds are still in the school system. But the target population was 4% smaller than the age cohort in the UK, 5% smaller in Austria, and 15% smaller in Korea. Only 50% of Mexican 15-year-olds were still in school. Young people leaving school before the age of 15 tend to be at the lower end or at very bottom of performance scales. Countries with high inclusion rates are thus likely to be penalized in terms of mean

¹ There can be few areas in which it is equally tempting for scientists to confuse science and politics. Educational economists, in particular, would be well advised to recall Max Weber's lecture *Wissenschaft als Beruf* [*Science as a Vocation*].

performance scores. The OECD publications fail to discuss the implications of these differences between age cohort and target population.

The OECD studies assume that the items assessing mathematics and science, but also reading, are culturally equivalent. Although considerable efforts were made to minimize cultural bias, test items have been found to show differential item functioning depending on the source language and the country concerned (Artelt, & Baumert, 2004). For example, French students have been found to be at a systematic disadvantage in items translated from English. It has not been possible to estimate the size of the bias against the East Asian countries, because these countries did not contribute sufficient test tasks to the item pool.

The provision of coaching measures intended to improve student performance on international comparison studies can be expected to vary from country to country depending on public interest in the studies' results. In an experimental study, Brunner, Artelt, Krauss, and Baumert (2007) found that it is possible to improve students' performance on the PISA mathematics test within a reasonable timeframe by means of coaching; it seems to be much more difficult to provide effective coaching for the reading test, however. In both cases, high-achieving students benefit more from coaching than their lower-achieving classmates. Intensive coaching can thus distort the empirical distribution of student performance.

What does the unidimensionality assumption of the underlying Rasch model actually mean? One can take a fundamentalist approach to this question by using selected sample items to show problems of fit and then condemning the test outright – as has been done by the group headed by Jahnke in Potsdam (Jahnke, & Meyerhöfer, 2006). Alternatively, one can take a rather more level-headed and empirically grounded approach to the constraints of unidimensionality assumptions, as has been done by Goldstein (2004).

In PISA, as in TIMSS, differential item functioning analyses suggest that the tests are in fact multidimensional, and that theoretically and practically meaningful differences in country profiles may be concealed if test instruments are designed to be unidimensional. For the TIMSS science test, for example, we found substantially different traditions of theory and teaching practice in the college-bound tracks in Sweden, France, and Germany, leading to specific performance profiles (Baumert, Bos, & Watermann, 2000; Baumert, Klieme, & Watermann, 1998). Goldstein (2004) asked whether it should not in fact be the task of international comparisons to highlight such differences in profiles. This is not possible with unidimensional tests.

Validity studies for all large-scale assessments are scarce. This applies to both construct validity and predictive validity. There may be much scope for criticism here; a recent study by Detlef Rost may be the tip of the iceberg (Rost, & Sparfeldt, 2007). Rost presented students with PISA reading items based on a certain stimulus text and systematically shortened that text until only the questions were left. It is a matter of some concern that some PISA test items can evidently be answered correctly without the stimulus text. What do these items measure?

The competencies measured in PISA are widely assumed to be indicators of human capital on the individual and societal level (OECD, 2000). Yet there are virtually no studies on the predictive validity of the tests. What is it that determines

a student's future career path: the competencies assessed in PISA – and if so, which? – or the formal qualifications he or she is awarded? To date, very few studies have addressed these kinds of questions. A rare exception was a recent U.S. study showing that formal qualifications seem to be decisive for the transition to the world of labor, but that once young people have entered the workplace, domain-specific competencies seem to be decisive for occupational success. There are no comparable studies for international large-scale assessments.

These weaknesses are not severe enough to justify blanket opposition to international educational assessments, such as continues to exist in Germany. Nevertheless, the OECD would be well advised to take a proactive approach to addressing these weaknesses, if only to avoid providing critics with ammunition.

5 Structural Limitations and Weaknesses of Cross-Sectional International Comparisons: The Explanatory Gap

The latest generation of international educational assessments provides a reliable descriptive framework in terms of the levels of student knowledge and skills, and how these are distributed across the student population. Their weaknesses become apparent when attempts are made to explain cross-country differences or to identify potential points of intervention for policy measures. Indeed, there are practically no consistent and robust findings of bivariate relations between characteristics of education systems and student performance levels, let alone practicable multivariate analyses of such relationships. In fact, the data available are best suited to put explanatory myths to the test, as Baker, LeTendre, and Boe have done for the alleged negative effects of the “broken curriculum” in the United States and for the supposed specific professional beliefs of Japanese teachers (Baker, & LeTendre, 2005; Boe et al., 2001). To some extent, the cross-sectional design of PISA and other large-scale international assessments is at odds with the study's objective of informing evidence-based policy. In fact, the OECD often gives the impression that causal inferences can be drawn from observational data, which puts it at great risk of confusing science and politics (OECD, 2007).

Attempts to explain achievement differences or differences in other performance measures among countries regularly come up against the limits of theoretical modeling, as illustrated by the following three examples. First, when aggregated to a higher level – be it the school or even the country level – variables often undergo a shift in meaning that is not immediately apparent. The only unmistakable indication that an aggregation shift has occurred is when the same variable has a different sign at different levels of aggregation. For instance, at the individual level, there is a robust relationship of intermediate magnitude between the self-concept of academic ability and students' performance status and performance gains. This relationship has been replicated hundreds of times in motivational psychology. When individual measures of self-concept are aggregated to produce a national index of self-confidence, however, there is a close negative relationship between self-confidence and a country's mean level of performance. The explanation for these findings is as follows: When aggregated to the country level, self-concept becomes an indicator of cultural differences in modesty rules.

Students in the East Asian countries are systematically more reserved in their self-attributions of proficiency. But the scales still differentiate between individuals within these countries to a similar extent – though at a lower overall level. This phenomenon, which can be termed aggregation bias, is often overlooked (OECD, 2004).

A second and even more widespread problem is inadequate model specification at the aggregate level. On the between-country level, the bivariate relationship between an education system's degree of decentralization on the operative level and its mean performance is weak but positive, suggesting that countries with a decentralized responsibility structure are more effective. Given that within countries the relationship between operative decentralization and the performance level of a school is negative, however, interpreting the bivariate correlation at country level without first testing it in a theory-driven multilevel model would be jumping to conclusions.

The third example relates to the ill-considered use of methodologically advanced reanalyses of international large-scale assessment data for political ends. Hanushek and Woessmann recently attempted to show that school systems that track students to different ability levels at an early age lead to increased variation in performance levels and thus exacerbate social differences (Hanushek, & Woessmann, 2005). Tracking is seen as a causal factor responsible for undesired effects. Without question, numerous descriptive findings indicate that social disparities increase when tracking is introduced. These findings make intuitive sense. However, surprisingly little is known about the mechanisms behind these increasing disparities.

Hanushek and Woessmann used the difference-in-differences principle to support what they suppose to be causal conclusions. They investigate the extent to which between-country performance differences established at the elementary level change when ability tracking is introduced at secondary level. Theirs is a strong model for analysis in which the frame of reference in each country is kept constant on the formal level. However, it would be incorrect to assume that educational systems of a country are homogeneous at primary and secondary level. Germany is a prime example here. Elementary schooling in Germany has always been based on a single conceptual tradition, with all elementary teachers receiving the same training. The same does not apply at secondary level. Rather, the current secondary tracking system perpetuates the 19th century tradition of higher and lower levels of schooling, even though timetables and curricula have been harmonized. In fact, these traditions are institutionally perpetuated in the different forms of training provided for teachers in the different tracks. As long as we remain unable to disentangle the effects of educational traditions, teacher training, and institutional tracking, we will remain unable to draw causal conclusions. As plausible, and welcome from a normative perspective, as such conclusions may be, they are politics and not science.

6 The Path to Excellence in Teaching and Assessment: Compensation of Weaknesses

Is it possible to compensate for the explanatory weaknesses of international educational assessments? One approach would be to incorporate the merits of cross-sectional surveys within a broader research program. A potential focus of this kind of extended research program might be on teaching, learning, and student progress. To this end, cross-sectional designs would have to be expanded to longitudinal studies with at least one, and preferably two, repeated measures. This would enable researchers to examine instruction from the perspectives of both teaching and learning, and would open up entirely new potentials for analysis, as illustrated by the following two examples.

The IEA's Third International Mathematics and Science Study was the first to implement, in a supplementary study conducted in the USA, Japan, and Germany, a video component which, 5 years later, was replicated in a broader selection of countries (Baumert, Lehmann, Lehrke, Schmitz, & Clausen, 1997; Stigler, Gonzales, Kawanaka, Knoll, & Serrano, 1999). The study investigated 8th-grade mathematics instruction, with one lesson per teacher being videotaped, resulting in samples comprising 50 to 150 lesson hours. In Germany, two further lessons were videotaped, thus allowing for an assessment of the transsituational stability of teachers' instructional behavior. Moreover, the TIMSS Study in Germany was extended to a longitudinal design embedding the video component between two measurement points. Thus, student learning gains could be related to instructional processes. Teacher ratings, student reports, and independent observations of instructional processes were systematically combined, and the specific validity of the three perspectives was examined (Clausen, 2002; Kunter, 2005).

The major finding of the video study was that – despite the variations observed among teachers within a country – there seem to be specific modal cultural scripts of mathematics instruction in each of the three countries. But these choreographic differences reflect only the visible surface structure of instruction, which – as the TIMSS replication study showed – is not related to performance outcomes. When analyses were focused on the mathematical and cognitive demands of the tasks presented, and on the mathematical challenge involved in these tasks and in classroom discourse, however, it emerged that – regardless of the social choreography of instruction and the manifest student activities (hands-on work, group work, etc.) – different patterns of cognitive activation could be discerned as a function of the task structure and of the mathematical involvement of students as triggered by classroom discourse (Knoll, 2003; Neubrand, 2002).

In all three countries, there is considerable variation in the form and the quality of mathematical learning opportunities. Nevertheless, some characteristic differences can be ascertained. In Germany, there is a widespread pattern in classroom discourse that can be described as a step-by-step trivialization of complex mathematical problems (Klieme, Schümer, & Knoll, 2001). Starting instruction with a challenging mathematics problem, teachers apply the Socratic maieutic method, but then trivialize the problem by progressing in small incremental steps, reducing the cognitive challenge of the discourse. In Japan, processes of cognitive activation are generated beneath the choreographic surface

structure of teacher-directed instruction – with alternating periods of teacher presentation and work in small groups – by the self-differentiating and self-individualizing potential of tasks with multiple solution paths. Classroom discussion on different strategies for solving these tasks achieves a high level of mathematical explication and cognitive demand. An instructional pattern frequently observed in the United States involves teacher presentation of a mathematical problem and the algorithm needed to solve it, followed by group work or seatwork on a single solution path. The mathematical challenge is constantly rather low. Results of the longitudinal study carried out in Germany showed that neither student involvement in the structure and planning of lessons, nor a hands-on approach to mathematics problems, nor the sharing of ideas in a learning community as such are responsible for successful instructional outcomes. What is crucial is the degree of cognitive activation and mathematical involvement, largely irrespective of the social setting of the classroom.

This finding is a direct extrapolation from the results of qualitative studies on mathematics instruction conducted in the United States by scholars such as Leinhardt, Borko, Schoenfeld, Ball, and Ma (Ball, Lubienski, & Mewborn, 2001; Borko et al., 1992; Leinhardt, & Smith, 1985; Ma, 1999; Schoenfeld, Minstrell, & van Zee, 2000). These studies have consistently shown the decisive impact of teachers' ability to explain and represent mathematical content at specific points during instruction that are critical for students' mathematical understanding. Interview studies and lesson transcripts suggest that the pedagogical content knowledge and representational resources available to teachers are limited by their command of content knowledge. A deep-level knowledge of the mathematical content taught at school seems to be a necessary prerequisite for the use of flexible mathematical knowledge for teaching.

This finding leads up to my second example. In the second cycle of PISA in Germany, PISA 2003, the cross-sectional design was again extended to a longitudinal one, allowing for a matching, and joint analysis, of learning groups and teachers (Baumert et al., 2004; Kunter et al., 2006). The focus of this research program was on teacher expertise. In contrast to the distal indicators of teacher knowledge that are often used – e.g., years of education or number of subject matter courses taken at university level or certification – we attempted to assess teachers' content knowledge – their understanding of mathematics as taught at secondary level – and their pedagogical content knowledge – the knowledge and the skills needed to make mathematics accessible to students – directly and separately. A test was constructed to tap teachers' deep-level mathematical understanding of the content taught at middle school or high school (Krauss et al., in press). Furthermore, a computer-based test was constructed to tap mathematics teachers' pedagogical content knowledge on three dimensions, namely:

- evaluating the instructional potential of tasks and optimizing the orchestration of tasks during instruction;
- rapid and accurate diagnosis of students' mathematical strategies, ideas, and errors and making productive use of them for fostering mathematical understanding;
- teachers' repertoires of different ways of explaining and presenting the same mathematical content.

The findings of the longitudinal study were unequivocal, providing compelling confirmation of the findings of the qualitative case studies conducted in the United States. Pedagogical content knowledge is highly predictive of mathematics teachers' epistemological beliefs and instructional goals, their selection of mathematically challenging and varied tasks for lessons and assessment, and the level of cognitive activation in classroom discourse. Mediated by these characteristics, pedagogical content knowledge is responsible for students' performance gains over the academic year. The net effect of pedagogical content knowledge is substantial (Baumert et al., submitted).

Moreover, pedagogical content knowledge proved to be directly related to a deep-level mathematical understanding of the content taught. Content knowledge is a necessary, but not a sufficient condition for cognitively activating instruction. Insufficient content knowledge limits pedagogical content knowledge and, hence, the repertoire of activities available to teachers in the classroom.

This research would not have been possible without conceptual groundwork and qualitative analyses. As I see it, this is a good example of how qualitative research, large-scale assessments, and research programs that extend and expand on these assessments can be combined to promote excellence in teaching and assessment.

References

- American Association for the Advancement of Science. (1998). *Blueprints for reform – Project 2061*. New York: Oxford University Press.
- Artelt, C., Schiefele, U., & Schneider, W. (2001). Predictors of reading literacy. *European Journal of Psychology of Education, 16*(3), 363-383.
- Artelt, C., & Baumert, J. (2004). Zur Vergleichbarkeit von Schülerleistungen bei Leseaufgaben unterschiedlichen sprachlichen Ursprungs [Comparability of students' reading literacy performance measured with items originating from different language backgrounds]. *Zeitschrift für Pädagogische Psychologie, 18*(4), 171-185.
- Artelt, C., & Schlagmüller, M. (2004). Der Umgang mit literarischen Texten als Teilkompetenz im Lesen? Dimensionsanalysen und Ländervergleiche [Understanding of literary texts as a subcomponent of reading literacy? Dimensional analyses and cross-national comparisons]. In U. Schiefele, C. Artelt, W. Schneider, & P. Stanat (Eds.), *Struktur, Entwicklung und Förderung von Lesekompetenz - Vertiefende Analysen im Rahmen von PISA 2000* (pp. 169-196). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Baker, D. P., & LeTendre, G. K. (2005). *National differences, global similarities. World culture and the future of schooling*. Stanford, CA: Stanford University Press
- Ball, D. L., Lubienski, S. T., & Mewborn, D. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), *Handbook of research on teaching* (pp. 433-456). New York, Macmillan.
- Baumert, J., Bos, W., & Watermann, R. (2000). Fachleistungen im voruniversitären Mathematik- und Physikunterricht im internationalen Vergleich [Student achievement in pre-college mathematics and physics in international comparison]. In J. Baumert, W. Bos, & R. Lehmann (Eds.), *TIMSS/III: Dritte Internationale Mathematik- und Naturwissenschaftstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn: Vol. 2. Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe* (pp. 129-180). Opladen: Leske + Budrich.
- Baumert, J., Klieme, E., & Watermann, R. (1998). Jenseits von Gesamttest- und Untertestwerten: Analyse differentieller Itemfunktionen am Beispiel des mathematischen Grundbildungstests der Dritten Internationalen Mathematik- und Naturwissen-

- schaftsstudie der IEA (TIMSS) [Beyond total test scores and subtest scores: Analysis of differential item functioning using the example of the TIMSS mathematics literacy test]. In H.-J. Herber, & F. Hofmann (Eds.), *Schulpädagogik und Lehrerbildung* (pp. 301-324). Innsbruck: Studien-Verlag.
- Baumert, J., Kunter, M., Brunner, M., Krauss, S., Blum, W., & Neubrand, M. (2004). Mathematikunterricht aus Sicht der PISA-Schülerinnen und -Schüler und ihrer Lehrkräfte [Mathematics instruction from the perspective of the PISA students and their teachers]. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost, & U. Schiefele (Eds.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (pp. 314-354). Münster: Waxmann.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Dubberke, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., & Tsai Y. (submitted). *Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress*.
- Baumert, J., Lehmann, R., Lehrke, M., Schmitz, B., & Clausen, M. (1997). *TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich Deskriptive Befunde* [TIMSS – Mathematics and science instruction in international comparison. Descriptive findings]. Opladen: Leske + Budrich.
- Baumert, J., & Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb [Family background, educational participation, and knowledge acquisition]. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann, & M. Weiß (Eds.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (pp. 323-407). Opladen: Leske + Budrich.
- Boe, E. E., Barkanic, G., Shin, S., May, H., Leow, C. S., Singleton, J. C., et al. (2001). *Correlates of national differences in mathematics and science achievement: Evidence from TIMSS* (Data Analysis Report No. 2001-DAR1). Philadelphia: University of Pennsylvania, Graduate School of Education, Center for Research and Evaluation in Social Policy.
- Borko, H., M., Eisenhart, M., Brown, C. A., Underhill, R. G., Jones, D., & Agard, P. C. (1992). Learning to teach hard mathematics: Do novice teachers and their instructors give up too easily? *Journal for Research in Mathematics Education*, 23(3), 194-222.
- Brunner, M., Artelt, C., Krauss, S., & Baumert, J. (2007). Coaching for the PISA test. *Learning and Instruction*, 17, 111-122.
- Clausen, M. (2002). *Unterrichtsqualität – Eine Frage der Perspektive?* [Instructional quality – A question of perspective?]. Münster: Waxmann.
- Fischer, G. H., & Molenaar, I. W. (1995). *Rasch models – Foundations, recent developments, and applications*. New York: Springer.
- Goldstein, H. (2004). International comparisons of student attainment: Some issues arising from the PISA study. *Assessment in Education*, 11, 319-330.
- Hanushek, E. A., & Woessmann, L. (2005). *Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries* (NBER Working Paper 11124). Cambridge, MA: National Bureau of Economic Research.
- Jahnke, T., & Meyerhöfer, W. (2006). *PISA & Co – Kritik eines Programms* [PISA & Co – Criticism of a program]. Hildesheim: Franzbecker Verlag.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.
- Kirsch, I. S., Jungeblut, A., & Mosenthal, P. B. (1998). The measurement of adult literacy. In T. S. Murray, I. S. Kirsch, & L. B. Jenkins (Eds.), *Adult literacy in OECD countries: Technical report on the first international adult literacy survey* (NCES Publication No. 98-053) (pp. 105-134). Washington, DC: US Department of Education, National Center for Education Statistics.

- Klieme, E., Schümer, G., & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: "Aufgabenkultur" und Unterrichtsgestaltung im internationalen Vergleich [Mathematics instruction at lower secondary level: "Task culture" and lesson structure in international comparison]. In E. Klieme, & J. Baumert (Ed.), *TIMSS – Impulse für Schule und Unterricht. Forschungsbefunde, Reforminitiativen, Praxisberichte und Video-Dokumente* (pp. 43-57). Bonn: Bundesministerium für Bildung und Forschung.
- Knoll, S. (2003). *Verwendung von Aufgaben in Einführungsphasen des Mathematikunterrichts* [Use of tasks in the introductory phases of mathematics instruction]. Marburg: Tectum Verlag.
- Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., et al. (in press). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology*.
- Kunter, M. (2005). *Multiple Ziele im Mathematikunterricht* [Multiple goals in mathematics instruction]. Münster: Waxmann.
- Kunter, M., Dubberke, T., Baumert, J., Blum, W., Brunner, M., Jordan, A., et al. (2006). Mathematikunterricht in den PISA-Klassen 2004: Rahmenbedingungen, Formen und Lehr-Lernprozesse [Mathematics instruction in the PISA 2004 classes: Structural conditions, forms, and processes of teaching and instruction]. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost, & U. Schiefele (Eds.), *PISA 2003: Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (pp. 161-194). Münster: Waxmann.
- Leinhardt, G., & Smith, D. A. (1985). Expertise in mathematics instruction: Subject matter knowledge. *Journal of Educational Psychology*, 77(3), 247-271.
- Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- McEneaney, E., & Meyer, J. (2000). The content of the curriculum: An institutionalist perspective. In M. Hallinan (Ed.), *Handbook of sociology of education* (pp. 189-211). New York: Kluwer/Academic Plenum Publishers
- Meyer, J. W., Ramirez, F. O., Rubinson, R., & Boli-Bennett, J. (1977). The world educational revolution, 1950-1970. *Sociology of Education* 50(4), 242-258.
- Meyer, J., Ramirez, F., & Soysal (1992). World expansion of mass education, 1870-1980. *Sociology of Education*, 65(2), 128-149.
- National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics*. Reston, VA: Library of Congress Cataloguing in Publication
- Neubrand, J. (2002). *Eine Klassifikation mathematischer Aufgaben zur Analyse von Unterrichtssituationen: Selbsttätiges Arbeiten in Schülerarbeitsphasen in den Stunden der TIMSS-Video-Studie* [An analysis of instructional situations based on the classification of mathematical tasks]. Hildesheim: Verlag Franzbecker.
- Organisation for Economic Co-Operation and Development (OECD). (1999). *Measuring student knowledge and skills. A framework for assessment*. Paris: Author.
- Organisation for Economic Co-Operation and Development (OECD). (2000). *Measuring student knowledge and skills. The Pisa 2000 assessment of reading, mathematical and scientific literacy*. Paris: Author.
- Organisation for Economic Co-Operation and Development (OECD). (2004). *Learning for tomorrow's world. First results from PISA 2003*. Paris: Author.
- Organisation for Economic Co-Operation and Development (OECD). (2007). *PISA 2006 Science competencies for tomorrow's world*. Paris: Author.
- Parsons, T. (1971). *The system of modern societies*. Englewood Cliffs: Prentice-Hall.
- Ramirez, F. O., & Boli, J. (1987). The political construction of mass schooling: European origins and worldwide institutionalization. *Sociology of Education*, 60, 2-17.

- Rost, D. H., & Sparfeldt, J. R. (2007). Leseverständnis ohne Lesen? Zur Konstruktvalidität von multiple-choice-Leseverständnistestaufgaben [Reading comprehension without reading? On the construct validity of multiple-choice reading comprehension items]. *Zeitschrift für Pädagogische Psychologie*, 21(3/4), 305-314.
- Schoenfeld, A. H., Minstrell, J., & van Zee, E. (2000). The detailed analysis of an established teacher carrying out a non-traditional lesson. *Journal of Mathematical Behavior*, 18(3), 281-325.
- Stanat, P., & Christensen, G. (2006). *Where immigrant students succeed – A comparative review of performance and engagement in PISA 2003*. Paris: OECD.
- 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. A research and development report*. Washington: National Center for Education Statistics.
- Walter, O., & Taskinen, P. (2007). Kompetenzen und Bildungsrelevante Einstellungen von Jugendlichen mit Migrationshintergrund in Deutschland: Ein Vergleich mit ausgewählten OECD-Staaten. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme, & R. Pekrun (Eds.), *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* (pp. 337-366). Münster: Waxmann.
- Wiseman, A., & Baker, D. (2005). The worldwide explosion of internationalized education policy. In D. Baker, & A. Wiseman (Eds.), *Global trends in educational policy. International perspectives on education and Society*. Amsterdam, Netherlands: Elsevier/JAI.