Measurement of vocational competencies - a contribution to an international large-scale assessment on vocational education and training[1]

Esther Winther[a*], Frank Achtenhagen[b]

[a,b]*Institute for Human Resource and Business Education, Georg August University, Göttingen*

Abstract

Measurement of vocational competencies which includes not only the knowledge but also the action domain is one of the central targets of VET research - under scientific as well political perspectives. It also is a decisive prerequisite for running an international large-scale assessment on VET. The article describes how such measurement is operationalized by defining dimensions of VET competencies and developing adequate items which mirror authentically tasks to be fulfilled at the workplace. Tasks and items are presented in the context of a technology-based virtual enterprise focusing on the fields of business and administration. Their format allows Item Response Theory (IRT) scaling.

Keywords: vocational competence, business education, authentic measurement of work-related tasks, IRT scaling

1. Vocational 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,

---

[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.

*Corresponding author: Platz der Göttinger Sieben 5, D-37073 Göttingen. E-mail: ewinthe@uni-goettingen.de.

- how to reform and adapt the given system of vocational 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 articled 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.

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 lifes, as well as to develop them to attentive, critical and constructive citizens, actively participating in social and political processes (Lang-von Wins, 2003). 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. 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). 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 its national differentiations. 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, there is a chance to get more specific knowledge about 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. But 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 by vocational education and training.

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?

For space reasons, only the first of 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 (Baethge et al., 2006).

We are convinced that 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 eighth-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; cf. Achtenhagen & Grubb, 2001). 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: an understanding of the dynamics of economic reality that includes a series of acts and decisions. Regarding the work of Gelman & Greeno (1989) we characterize the understanding of real-world dynamics as a set of knowledge structures that imply the kinds of successful performances of the learners.

   - Static-linearized points of views of operational reality have to be given up and a system-oriented approach to business studies should be pursued. System-orientation means to expose the connections between different topics, concepts and instruments and to integrate the different elements

of knowledge into a meaningful whole by which the manifold specific and dynamic relations are open for observation and treatment. This systemic view on business studies has to be complemented by a process-orientation. Measurement has necessarily to consider this systemic view (Winther & Achtenhagen, 2008).

- Recent research and development shows that the link between cognitive goals and knowledge to self-related functions gains more and more importance (Anderson & Krathwohl, 2001).

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 generate 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. Recent approaches are the above mentioned development of a European Qualifications Framework (EQF) or a European Credit Transfer System for Vocational Education and Training (ECVET). 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 by simultaneously considering aspects of assessment and research on learning and instruction.

In the following it will be shown step by step how such construct can be theoretically and practically developed, how corresponding test items can be formulated and presented in a technology-based way.

4.   Vocational competence: dimensions of complexity and tasks

Taking into account these necessities the measurement of vocational competence needs a comprehensive and multidimensional construct (see Figure 1).

A VET study must focus on at least three different *competence levels* and their interrelationships (see Greeno et al., 1984):

- conceptual competence corresponds to factual knowledge as knowledge of facts, structures, and knowledge nets that can be transmitted into action schemata;

- procedural competence subsumes the application of knowledge, i.e. how to operate with facts, structures, knowledge nets and their corresponding elements;

- interpretative competence provides connections between features of the problem setting and goals of the learner.

All in one these competence levels refer to successfully carried out occupation-specific tasks (the necessary skills contain occupation-specific content, job-specific as well as enterprise-specific skills - see Oates, 2004).

Each level of competence corresponds with *different dimensions of complexity*. Is an occupation-specific task familiar to the learner, is that task practised in school and work or simply structured that many learners can be successful in carrying out that task? The use of different dimensions of complexity as a heuristic reference model assists researchers to identify specific hurdles in learning and performance. Furthermore, the use of different dimensions of complexity provides indications of how a task should be designed for increasing complexity and, thus, for more or less chances of success.

- *Modeling* refers to learners' understanding and learning processes with regard to performance situations in specific settings in school and at the workplace - in the tradition of applied cognitive psychology (see Glaser, 1976; Shavelson et al., 1992; Yuan et al., 2006). Contrasting high achieving and low achieving learners the modeling category is able to provide an insight into learners' ability to develop and set concrete models corresponding to real-world challenges and tasks and to activate models for solving that challenges and tasks.

- The *cognition* category relates to Bloom's taxonomy of cognitive processes and Anderson & Krathwohl's revision (Anderson & Krathwohl, 2002). Learners' levels of success correspond with the cognitive process dimension necessary to carry out an occupation-specific task.

- *Content* codes categories of curriculum. This dimension covers both learners' knowledge base and their realms of experience. In addition learning and coping strategies come into play. Recent research and development shows that the link between cognitive goals and knowledge to strategies of learning and performance gains more and more importance (Anderson & Krathwohl, 2001).
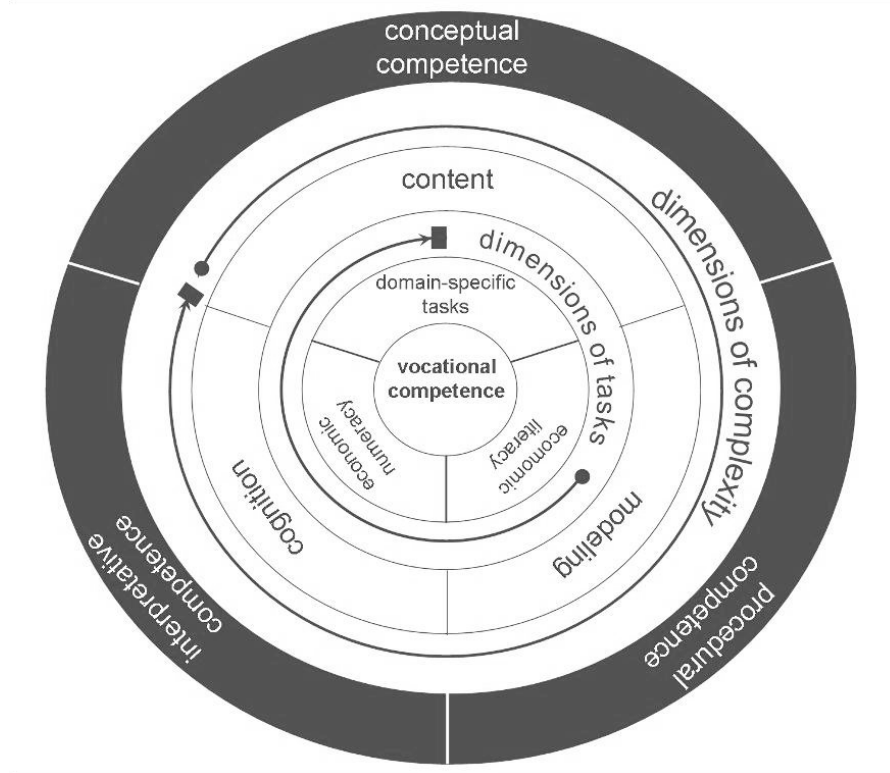
Figure 1: Construct of vocational competence

The performance of the learners can be standardized by such reference models, so that existing deficits while carrying out the tasks can be determined more easily. The dimensions of task focus on labour market requirements:

- On the level of *generic occupational* requirements, tasks refer to the collection of required individual prerequisites in order to successfully develop occupational identities and to solve work-related tasks. Such prerequisites are economic literacy and numeracy. In the tradition of Gelman & Greeno (1989) we interpret both dimensions as *domain-linked*. The domain-linked category refers to the notion of occupational key skills, which comprise knowledge about specific methods and strategies (Rychen & Salganik, 2003). Concepts like literacy and numeracy have been discussed for a long time in the context of international comparative studies of compulsory education (OECD, 1995; 2003).

- *Domain-specific* tasks take into consideration *specific occupational* require-

ments. These requirements contain occupation-specific contents, job-specific as well as enterprise- specific skills (see Oates, 2004).

For assessing the coping with such tasks we recommend to design them as part of technology-based test situations (Achtenhagen, 2004). The area of content - generic as well as specific occupational requirements - provides the base for an adequate understanding of a company with its systemic economic coherences. It is supposed to give the learners firstly an overview and to enable them to classify and to structure the multiple details and facts of a really existing company meaningfully. The purposes and aims of a company are discussed against the background of different groups of demands and interests, for example of stakeholders or shareholders. Thereby, it becomes clear for the learners that companies are living and mouldable economic and social systems that have to be analyzed in their relations within the market. Therefore, different technology-based tools were developed: storyboards of authentic business processes, applications of integrated ERP-systems including customers and suppliers list, in-depth analyses of the designed company, and so on (for more information see item part).

5.   The corresponding item-response construct map

The construct "vocational competence" is mainly defined by a set of items (tasks on general and specific occupational requirements) and the responses to those items (ratings on a heuristic reference model). For practical research reasons it is easier to start by ordering the item responses (Wilson, 2005; 2008). A sketch of the construct map for the construct "vocational competence" is shown in Figure 2. The item-response construct map indicates the qualitative differences in item responses using the heuristic reference model. That framework demonstrates the qualitative structure of levels inherent in the construct referring to different complexity levels. The responses to the items are also able to contrast response patterns/ behaviour that vary from low to high complexity:

- The *modeling* dimension extends from incomplete to complete modeling. Its complexity depends on (1) the level of familiarity, (2) the level of practice, and (3) structured hints given in occupation-specific settings. Whereas the levels of familiarity and practice are strongly related to personal conditions, structured hints are affected by the measurer. Reducing the hints in a specific setting increases the complexity of the setting. The learners have to build a model of the task using their cognition system. Expanding the hints improves the access. The score level of learners' responses extends from incorrect modelling (level 1) to advanced modelling (level 4). Whereas an incorrect model contains a reference to some appropriate terms or concepts without any understanding of the system and its coherences, the advanced model includes all expected relevant terms and concepts and goes also beyond them.
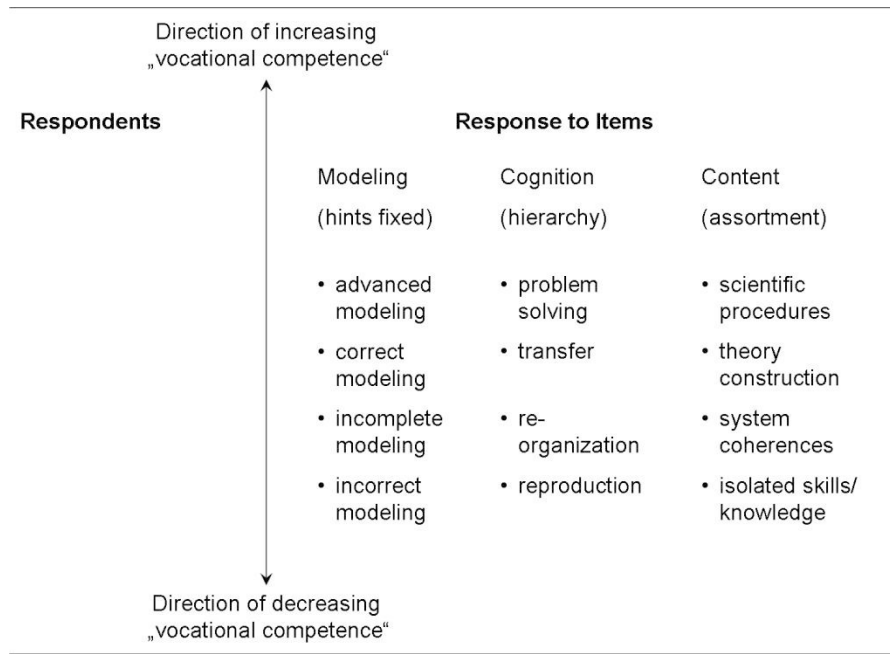
Direction of increasing
„vocational competence"

Respondents   **Response to Items**

Modeling	Cognition	Content
(hints fixed)	(hierarchy)	(assortment)
• advanced modeling	• problem solving	• scientific procedures
• correct modeling	• transfer	• theory construction
• incomplete modeling	• re-organization	• system coherences
• incorrect modeling	• reproduction	• isolated skills/ knowledge

Direction of decreasing
„vocational competence"

Figure 2: Construct map for "vocational competence"

- The *cognition* dimension ranges from reproduction to problem-solving. Learners' cognitive processes depend on both the understanding and the abilities to act. The major idea of the taxonomy is that what measurers want learners to know and they want to measure can be arranged in a hierarchy from less to more complex. Using the cognition category measurers are able to discriminate learners on following levels: On level 1, the reproductive level, a learner has simply the ability to describe and to replicate some characteristics of matter. Doing so, the learner is in the grip of the specific task. There are no signs of using other perspectives and ideas. When learners can begin to combine and interpret task-related facts and concepts in a more global view they are considered to have moved to level 2, re-organization. At the transfer level (level 3) learners apply all relevant facts and concepts. Moreover, they are able to transfer relevant task-related information to make their decision in an integrated and well-founded way. Level 4, problem-solving, involves goal-directed thinking and action in concrete occupation-specific situations. Learners self-organize their problem-solving process; that includes the selection of the best solution regarding upcoming economic developments and strategic decisions - also weighing intended main- and possible unintended side-effects.

- The *content* dimension varies from isolated skills and knowledge to scientific techniques and procedures. The measurer can select a representative assortment that includes content from low to high complexity. The content dimension in increasing order is stated as follows: The development, enlargement and deepening of basic economic skills and knowledge (level 1); the development, enlargement and deepening of the understanding of system coherences in the economy and basic principles of micro-economic and macro-economic behaviour considering also ecologically and socially relevant effects (level 2); the ability to have a systematic point of view on micro- and macro-economic coherences in exemplary representations as well as approaches of economic theory (level 3); the development, enlargement and deepening of fundamental scientific working techniques and procedures (level 4).

In addition to defining learners' responses at each level of the continua of "modelling", "cognition", and "content", notion of format and design of the test help to clarify the expected differences between levels; that also includes the notion of common errors in taking the test (Wilson, 2008). However, the presented construct map of vocational competence can serve as a framework for item design and a method of making measurement possible, and diagnoses interpretable.

6. Match between working processes and assessment

The construct map is item-response based. Assessment items, by contrast, are designed to stimulate respondents' behavior. Therefore, in the first place an analysis of possible respondents' behavior seems necessary: What kind of behavior is conventional in real-world processes and how can these modes of behaving be observed? These questions also include the description of real-world stimuli, labor conditions, tools and supplies for work as well as personal abilities. Keeping in mind those requirements the assessment style for measuring vocational competence can be defined: Based on the requirements of labor market, technology-based real-world items are used. There is a variety of performance criteria for technology-based test items. Technology-best test formats and executions reduce the influence of test environment and tester; therefore, an increase of test objectivity can be expected. In addition, authentic test situations take into consideration the situational character of competencies; as a result an increase of test validity follows. Furthermore, the impact on test motivation is obvious. Embedded multi-channel and interactive test stimuli given by video, animation or authentic tools are able to enlarge the test persons' effort and attendance in filling in the test.

All items are based on real-world working and business processes and can be defined as "challenges" in authentic situations. The idea behind the use of authentic challenges as assessment style is to observe authentic behavior in authentic test environments as key indicators for vocational competence in real-world environments. Translated to classroom assessment that kind of assessment is called "embedded"

(Black et al., 2003). As in classroom situation assessment should be an element of classroom teaching and learning processes assessment in vocational situations becomes part of the employees' working and developing processes. Using the Black et al. metaphor, we can think of it as being assessment "for" working. Therefore, the design of authentic items is based on three separate but dependent models:

1. a model of a real-world labor market is used to display interactive processes between individuals and institutions;

2. a model of real-world business processes presents concrete business processes in a complex way and describes procedures and workflows in different departments and between them;

3. a model of real-world business decisions focuses by using practical work examples on concrete decision-making processes to measure operation flexibility.

According to these models two dimensions of items are useful to create: First of all, items that measure an adequate understanding of procedures in working life: "domain-linked" items; for these items the concepts of economic literacy and economic numeracy are inherited. Economic literacy items correspond to knowledge of text and concepts including graphical elements and the understanding of quantitative details in the occupational field of business and administration. Economic numeracy subsumes the application of mathematical understanding, i. e. how to operate in an informed way with concrete economic data and processes but also to interpret adequately real business data.

In addition and consequently focused on working life requirements in specific occupations items are designed that stimulate an adequate operating in company departments with their specific economic coherences. Items of this category are called "domain-specific challenges". Each domain-specific challenge refers to a systemic view of business data and procedures with respect to economic decision-making processes - here in the fields of business and administration.

7.   Examples for domain-linked and domain-specific items

In a first study domain-linked data were collected using a paper-pencil test. Current work includes the integration of both domain-linked and domain-specific items in embedded technology-based assessment. Domain-linked items are basically focused on reviewing files, records, and other documents to obtain information to respond to requests (Winther, 2006).

Examples for domain-linked items in a paper-pencil test (translation from German) The testees get a complete article of a newspaper which contains a description of an economic situation and uses also graphics.

(6) You are a stockholder of Volkswagen AG. On December 31, 2003 the stock price is 17,500 Euro. Please, use the graphic to answer the following questions. (ORIGINAL: Sie sind Aktionär der Volkswagen AG. Der Wert Ihrer Volkswagen-Aktien betrug am 31.12.2003 17.500 Euro. Verwenden Sie für die folgenden Berechnungen bitte die Angaben aus der Grafik des Zeitungsartikels.)

...

(6c) Please calculate your lost in percent comparing it to an DAX 30 investment. (ORIGINAL: Im Folgenden berechnen Sie bitte Ihren Verlust in Prozent im Vergleich zu einer Investition in den "DAX 30".)

...

(6d) Please find arguments for stock price rise you can announce on ASM. (ORIGINAL: Sie sind über Ihren Verlust empört. Welche Maßnahmen, von denen Sie sich prinzipiell eine Kurserhöhung versprechen, könnten Sie als Aktionär von der Volkswagen AG fordern?)

Item (6c) counts to the economic numeracy block. Test persons have to calculate percentages of profit and loss regarding a stock investment. Item (6d) refers to the economic literacy block. Economic coherences, effects on the market, shareholder and stakeholder relationships but also individual options on the market are embedded in this task.

The items correspond with dimensions of complexity (for details see the construct map part). Both items differ in a considerable way:

- Regarding the *modeling dimension* the economic numeracy item calls for a basic understanding of percentage calculation: All hints for the calculation are given twice in graphical and in numerical form. In contrast, the economic numeracy item is much more complex. Test persons have to create a complete model of stock market and the coherences for the economy but also for individuals.

- Both items are stimulating different *cognitive processes*: Whereas the numeracy item refers to the re-organization level, the literacy item turns to the transfer level. The test persons have to consider all relevant facts that control the stock price.

- The *content dimension* varies too: The calculation example is focused on an isolated skill. On the other hand, the literacy example needs an economic theory-based construction in order to solve the item.

Examples for domain-specific items in technology-based assessment

The technology-based assessment is focused on the meaning, the range and the level of specific work-related skills and abilities in the occupational fields of business and administration. In order to provide authenticity, content validity and practical adaptivity during the test the assessment is grounded on an existent company as blue print. The modeling requirements are strict: First of all, a transparent method of production - together with a short parts list of production items - is necessary to ensure an understanding of company's objects and standings. Furthermore, well-known products are of interest to guarantee identifying marks during the test. Taking in mind those requirements the technology-based assessment is based upon a metal working industry. The company is called ALUSIM GmbH and fabricates packaging made of aluminium. Company's products are beverage cans (Coca Cola), cosmetics (NIVEA) and food grade packaging. Authentic parameters of modeling were used: a company history as well as company reporting like balance sheet, profit and loss statements and cash flow analyses. Furthermore, an authentic large-volume production and its corresponding costs were designed together with real-world customers and suppliers as well as their relationships. This includes the development of different technology-based tools like storyboards of authentic business processes, applications of integrated ERP-systems including customers and suppliers lists, in-depth analyses of the designed company, and so on.

All these products and modeling steps are necessary to stimulate performance in authentic vocational situations and work processes as key elements of competence measuring. In other words: A simulation of real-world conditions ensures a measurement of authentic abilities without bringing the testees in such real-world situations. Therefore, the business simulation contains complete business processes including reporting as well as running work processes corresponding to real-world activities. The focus on business and work processes guarantees the interpretability of test results. A booklet-design provides an acceptable testing time. Business

and work processes can be used in random order and separated in sequences of various interests: sales, purchases, and production challenges including different work processes for each of them. Furthermore, a simulation in contrast to real-world test environments assures a content-related and cognition-based characterization of test environment and provides a valid description of personal abilities and item difficulties with regard to concrete test challenges. Therefore, the simulation permits a psychometric modeling approach.

All domain-specific challenges are video-based and refer to a complete sequence of tasks (i. e. work processes). The test persons navigate through the tasks using an interactive desktop (Figure 3) which encloses all documents, ERP-systems, filings and helping tools the test person needs to fill in the tasks.
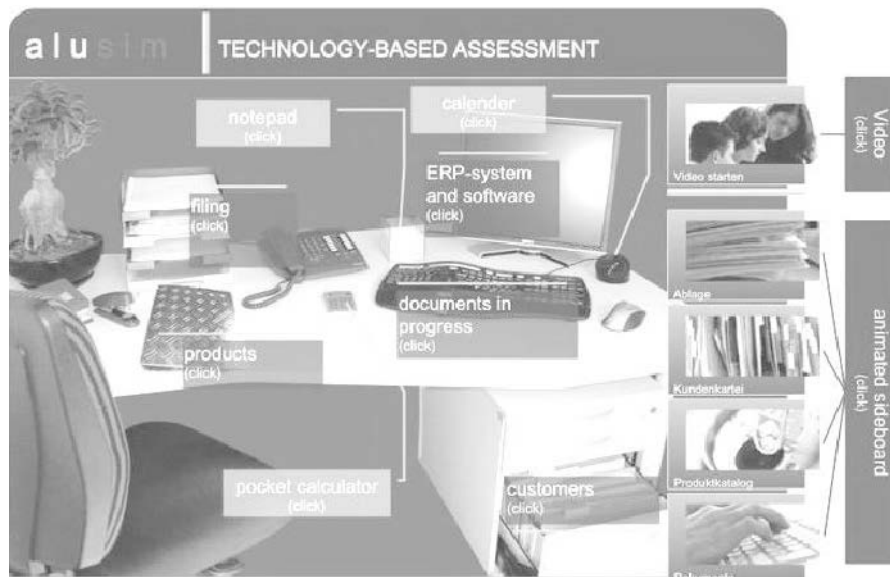

Figure 3: Interactive desktop technology-based assessment

A look behind the user interface clarifies how the simulation is working. Starting the video (click at the video button) opens the domain-specific challenge. The test persons watch a dialogue between a trainee and an employee. This dialogue contains useful information and a reference to the solving process. For example, in a sales challenge the trainee has to close a sale. With respect to the production process the test persons complete the ERP-system. The test persons analyze and monitor sales records, trends and economic conditions to anticipate consumers' buying patterns and determine what the company will sell and how much inventory is needed (Figure 4).
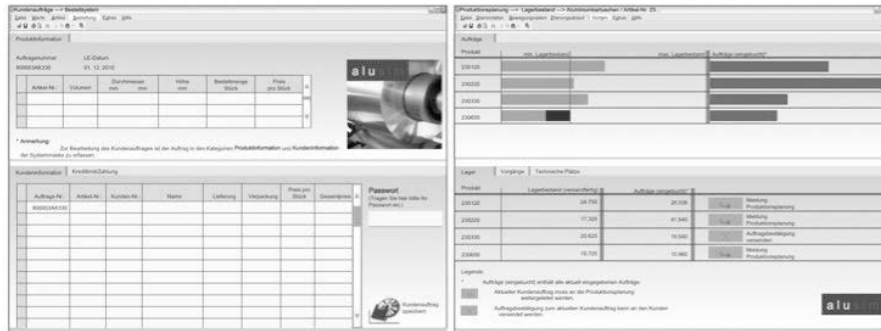
Figure 4: Example of technology-based assessment: sales challenge

Each challenge is structured by different work processes. The sales challenge, for example, includes two different steps. First, the trainee/test person has to enter customer data in an ERP-system. Therefore, he/she has to check out all relevant customer information, the ERP requirements, and the production process. In a second step, the trainee gets the information that the sale cannot be closed on usual terms. The test person has to manage that situation in contacting the customer. The underlined work process demands an deep understanding of sale records, economic conditions, customers' buying patterns and sensitivities. The test person has to anticipate the whole production process in order to make the right decision and to inform the customer correctly. It is obvious that both work processes refer to different complexity levels.

- Regarding the *modeling dimension* both the enter process and the react process do not differ. The work processes assume that a complete model of sales procedures, production, and customer services is known.

- With reference to the *cognition dimension* the first work process (enter process) refers to level 2, the re-organization level. Test persons have to combine and interpret task-related facts to solve the task. The second step (react process) stimulates the problem-solving level (level 4). Subject to trainees' experiences they have to self-organize an unusual situation; that includes the selection of the best solution regarding upcoming proceedings and economic developments with respect to company's objectives and customers' pretensions.

- The *content dimension* is not much different. The enter process in agreement with company´s production process is focused on the understanding of system coherences and the basic principles of ERP-systems. For that reason the work process is related to level 2, system coherences. The react process requires the ability to understand system coherences too. In addition, that work process

is also focused on the coherences between individual and institutional conditions of the market in a more micro-economic point of view (level 3, theory construction).

All items, domain-linked as well as domain-specific, are designed to generate diagnoses of test persons´ responses. Response behavior is coded with respect to the different complexity dimensions of the construct map and depends on the concrete item.

8.   Further steps of development and analysis

The test items developed on the basis of the technology-based virtual enterprise are tested with learners who run through the apprenticeship of "Industriekaufmann" (industrial clerk) in the German "Dual System" of VET. The overall evaluation of that test approach shall provide the reasons for the decision whether such item format might be a possible note to construct test items for vocational action competence within a large-scale assessment on VET.

The empirical collection of domain-specific data is ongoing. It has started in October 2008; so first results (based on IRT-models) of our study are expected in March 2009. According to TIMSS and PISA vocational domain-specific competencies will be reported mainly through scale scores derived using Item Response Theory (IRT) scaling. This approach allows the performance of a sample of test takers in a subject area to be summarized on a common scale or series of scales even when different persons have been administered different items. This also allows the performance of test takers from different cohorts to be summarized on a common scale even when they have worked on different test forms.

The international comparison of domain-specific competencies is new. That is, new tests for domain-specific competencies in the vocational area of interest have to be developed. The proposed item format is a realistic challenge in a computer-simulated work environment. The main advantage of this format is the (face) validity. That is, the test reflects professional tasks that are common in the vocational area. It ensures that the test measures the concept, which is intended to measure. Another advantage of the computer-simulated format is that students will enjoy doing the test, and this will enhance the data quality. The main disadvantage of the computer-simulated format is the relatively long assessment time. At least twice the time is needed to finish the test, in order to get the same measurement precision as with an abstract multiple-choice format for the items. Since the requirements on measurement error are relatively low, the test length will still fit in the time restrictions, which are reasonable in a large-scale assessment.

References

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*. Bonn, Berlin: BMBF.

Achtenhagen, F., Nijhof, J. W. & Raffe, D. (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, 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*. Washington, DC: American Educational Research Association.

Anderson, L. W. & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing. A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Baethge, M., Achtenhagen, F., Arends, L., Babic, E., Baethge-Kinsky, V. & Weber, S. (2006). *PISA-VET - A feasibility-study*. Stuttgart: Steiner.

Black, P., Harrison, C., Lee, C., Marshall, B. & William, D. (2003). *Assessment for learning*. London: Open University Press.

Buttler, F. (1992). Bildungslandschaft bis 2010. In F. Achtenhagen & E.G. John (Eds.) *Mehrdimensionale Lehr-Lern-Arrangements*. Wiesbaden: Gabler.

Europäische Kommission (2002). *Beschäftigung in Europa 2002 - Jüngste Tendenzen und Ausblick in die Zukunft*. Brussels: http://europa.eu.int/comm/employment_social/news/ 2002/sep/employment_in_europe2002_de.pdf.

Gelman, R. & Greeno, J. G. (1989). On the nature of competence: principles for understanding in a domain. In L. B. Resnick (Ed.) *Knowing, learning and instruction. Essays in honor of Robert Glaser*. Hillsdale, N. J.: Lawrence Erlbaum.

Glaser, R. (1976). Cognitive psychology and instructional design. In D. Klahr (Ed.) *Cognition and instruction*. Hillsdale, NJ: Lawrence Erlbaum.

Greeno, J. G., Riley, M. S. & Gelman, R. (1984). Conceptual competence and children's counting. *Cognitive Psychology*, 16, 94-143.

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*. Stuttgart: Schäffer-Poeschel.

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*, 2(1), 53-71.

OECD (1995). *Literacy, economy and society. Results of the first international adult literacy Survey*. Paris: OECD.

OECD (2003). *The PISA 2003 assessment framework - mathematics, reading, science and problem solving knowledge and skills*. Paris: OECD.

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*. Cambride, Göttingen: Hogrefe & Huber.

Shavelson, R. J., Baxter, G. P. & Pine, J. (1992). Performance assessments: Political rhetoric and measurement reality. *Educational Researcher*, 21(4), 22-27.

Wilson, M. (2005). *Constructing measures: An item-response modeling approach*. Mahwah, NJ: Erlbaum.

Wilson, M. (2008). Cognitive diagnosis using item response models. *Journal of Psychology*, 216(2), 74-88.

Winther, E. (2006). *Motivation in Lernprozessen. Konzepte in der Unterrichtspraxis von Wirtschaftsgymnasien*. Wiesbaden: DUV.

Winther, E. & Achtenhagen, F. (2008). Kompetenzstrukturmodell für die kaufmännische Bildung. Adaptierbare Forschungslinien und theoretische Ausgestaltung. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 104, 511-538.

Yuan, K., Steedle, J., Shavelson, R., Alonzo, A. & Oppezzo, M. (2006). Working memory, fluid intelligence, and science learning. *Educational Research Review*, 1, 83-98.