Oser, Fritz; Salzmann, Patrizia; Heinzer, Sarah

**Measuring the competence-quality of vocational teachers: An advocatory approach**

*Empirical research in vocational education and training  1 (2009) 1, S. 65-83*

in Kooperation mit / in cooperation with:

SKBF | CSRE

*http://www.skbf-csre.ch*

# Measuring the competence-quality of vocational teachers: An advocatory approach

Fritz Oser[a], Patrizia Salzmann[b], Sarah Heinzer[c,*]

[a,b,c] *University of Fribourg*

**Abstract**

Using a sample of 110 persons, this study examined the "advocatory approach" for predicting the quality of (VET) teacher's competency profiles. The participants (teachers, experts, non-teachers) observed a concrete teaching behaviour using a film vignette, and judged it by means of different quality criteria. The hypothesis that differences in the groups, who evaluate, elicit differing sensitivity about the quality of specified teaching competencies (and not teaching performance) was mostly confirmed. We suggest that teachers can individually compare their values with expert values and non-teacher values, and thus use the discrepancies in quality sensitivity for the purposes of self-guided in-service training.

*Keywords:* teaching, professional competence, competency profiles, standards, measurement, feedback

## 1. Introduction

### 1.1 Formulating competency profiles

The impact on the quality of instructional and vocational educational training (VET) of framing or formulating standards of teaching competence should not be underestimated. There is an impressive literature on the subject (e.g. Borich, 1977; Shulman, 1986; Helmke & Schrader, 1993; Weinert & Helmke, 1996; Beck & Dubs, 1998; Max, 1999; Oser, 2001; Shulman, 2007; Baumert & Kunter, 2006; Seidel & Shavelson, 2007; Beck, 2006; Nickolaus, 2007).

Competence as a notion includes more than just knowledge: it includes action, motivation, problem-solving and the learning context. When we say a vocational instructor has the competence to introduce an apprentice into how a complicated engine functions, we presuppose that the instructor is able to anticipate possible mistakes and failures by using different methods or information, differing time-sequences, different levels and stages of complexity, or to undertake different actions and provide different forms of feedback at each phase of the learning process. At

---

the same time, the apprentice develops new output-oriented competencies, namely to deal with the engine in an effective and functional way.

Why is it so difficult to formulate such teaching or training competencies? Why do we have so much trouble measuring them, and how can we diagnose their quality within the instructional process?

*First*, teaching competence represents only one part of the whole instructional process of a lesson. When we speak about group work undertaken by teacher X as one competency profile, we know that X uses different other competency profiles during the same lesson, e.g. direct teaching, questioning. - *Second*, teaching competence always represents a cluster of teaching actions. In the case of group work, these include introducing students to the work, structuring the learning parts, organizing and forming groups, explaining the task, distributing the material, scaffolding the interaction of individual students in the group, controlling the teamwork, incorporating the results of group work into the continuation of the lesson, and so forth. All these actions are part of one "multiply structured competence". This is why we speak of a *competency profile* and not of a single competence. Actions within such a competency profile are guided and directed by one and the same goal and occur in one and the same situation. In the case of group work, it is to cooperatively develop a new concept demanded by the curriculum. The term competency profile thus reflects "an impact of teaching actions in complex situations" (Oser et al., 2006, pp. 17f). They are formulated at a medium level of abstraction (micro formulations would concern single competencies, macro formulations the general tendencies of action clusters such as "to develop social competence"). - *Third*, a competency profile must be measurable and be measured with respect to quality. We can do this either if there are clear benchmarks for comparing bad and good teaching, or if we can compare the effort of a single teacher with the mean of a representative sample group of other teachers with similar tasks, or if we have expert teachers examine the implementation of a competency profile. Thus, the term professional standard can only be used if the condition for measurement is given.

A good definition of a professional teaching standard - one that goes beyond that offered by Frey (2006) or Weinert (2001) who speak in terms of general capacities - contains five elements. The first implies professional specificity. Here we address the professional competencies of vocational school teachers as a particular profession possessing specific professional knowledge (Bromme, 1992; Zeichner, 2006). The second element concerns specific actions. Competency profiles are implemented in class, and their quality only becomes apparent in the context of the situational accomplishment. The third element is related to the Taylor system and the holistic character of teaching. This refers to cutting the workflow (here: the flow of teaching) into distinct parts, each of which represents a different but similarly structured competency profile. The fourth is measurement. And the fifth is the ethical responsibility not to disadvantage students on the basis of their (class) backgrounds. Thus, competency profiles do not only imply theoretical, pedagogical, content-oriented, and

pedagogical content knowledge[1], but must enable vocational school teachers to act in everyday classroom interaction even under difficult conditions (see "Emergency Room Class"; Oser, 2001).

Within the project "Professional Minds", a research program dealing with what VET teachers do that leads to a high level of professionalism, we developed and validated diagnostic instruments to predict the quality of individual competency profiles among vocational teachers. "Diagnostic" means that researchers and teachers themselves both test and hypothesize about the value of specified quality criteria with respect to teacher behaviour, in contexts outside their real teaching.

During the first phase of the project, in a quasi-Delphi-study, 45 competency profiles were developed and validated. In order to understand what vocational school teachers really do and are concerned with, this Delphi approach started with a bottom-up procedure. In cooperation with experienced vocational school teachers, a very large number of competency profiles were formulated, and these were subsequently evaluated and selected. An example of a competency profile is: "The VET-teacher is able to connect the workplace-experiences of the apprentices with new content of his teaching, to model it and to develop with its help high new valid student-competencies." An other example - already mentioned in the paragraph upoff - would be: "The VET- teacher is able to organize different forms of group-work, to scaffold each single group, and to connect the group results into the continuous teaching." The validation of all the 45 standards was undertaken by a sample of 789 vocational school teachers, and the criteria encompassed importance, frequency, difficulty and significance for training.

Interestingly 44 of the 45 competency profiles were rated as quite important to very important (the last competency profile, which concerned the representation of the school at formal occasions, was rated as not very important). Frequency and significance with respect to the teacher training were found meaningful by the judges, especially as they were highly engaged in their own teaching, and therefore believed that these competencies ought to be introduced as part of the teachers' education. The difficulty of application was rated modestly, between quite not to quite difficult; extreme values did not appear.

In the next research phase, the prediction of a certain level of quality of teachers' competencies was an important concern. Using these 45 competency profiles, we developed a new "advocatory approach" for the diagnosis of competence. The example we employed was the competency profile entitled "giving feedback".

## 1.2    The "advocatory approach" as a new method of measuring competencies

The "advocatory approach" provides an indirect but highly generalizable prediction of how well a vocational teacher performs with respect to a certain competency pro-

---

[1]From the presentation "Zur Entwicklung und Situation der Berufsbildungsforschung im deutschsprachigen Raum" of J. Baumert on Februar 3, 2006 on the occasion of the resignation of Prof. Dr. Helmut Fend at the University of Zurich

file. Advocatory means here that the rating of a teacher says something about his/her own real teaching competencies (not performances). In general, two methods for measuring competence have been used: direct observation and self-evaluation (Frey, 2006). However, these two methods are not suited for measuring competency profiles in daily teaching. Direct observation demands high personnel and financial resources, and intervening variables such as social desirability are virtually impossible to control. Personal value judgments of teacher training coaches or in-service training coaches prevail. For its part, self-evaluation has insufficient validity or prognostic reliability, for rather than measuring competencies themselves, what is measured instead is whether a teacher estimates himself as competent (Fahrenberg et al., 2007; Frey & Balzer, 2005).

Due to these difficulties, we developed the alternate diagnostic method (Oser et al., 2007) shown in Figure 1. Person A (a novice, an experienced teacher or an expert) judges the teaching process of teacher B, and information from various such judges help to form the competency judgments by the research group. Teacher B is shown in a real classroom, in a film vignette stored on a computer disc, simultaneously from three perspectives: the entire classroom, the student to whom the teacher relates at a particular time, and the teacher. These three perspectives are evaluated using various categories (see the 'general criteria' and 'standard-specific criteria' boxes in Figure 1) with respect to a particular competency profile. A criterion-based questionnaire follows the presentation of the film vignette.

We can ask a variety of questions about the validity of this approach. What is the content if we, the researchers, have data on how a teacher judges another teacher? Do we know how they teach with respect to a real situation, or do we know something about their teaching performance? Can we say anything about the teaching quality of the teacher who observes and judges in general, or with respect to what they see in the film vignette?

These questions are of central importance, and we - the research group - take a very modest position. *First*, we believe the "advocatory approach" has many advantages. We can compare the judgements as to quality of different teachers with respect to the same situation, since the underlying material can be presented to a variety of judges. We can also use the same indicators of assessment repeatedly, which in the real-life lesson evaluation seldom exists or is only interpreted with respect to different individual lessons. *Second*, we are of the opinion that by their judgments, the observing teacher indicates his or her own competence - and not performance since competencies are preconditions for performance. Without knowing how a situation can be framed (professional knowledge and experience) we will never be able to make that situation function. *Third*, we believe teachers watching this film vignette and evaluating the observed teaching behaviour - able to replay a sequence as often as they like, stopping and investigating a detail, even correcting their own judgments - show a higher or lower level of professional awareness and sensitivity. Sensitivity is a professional precondition and a motor for change and it is a sign of differentiation among the respective evaluator groups (novice teachers, experienced teachers,
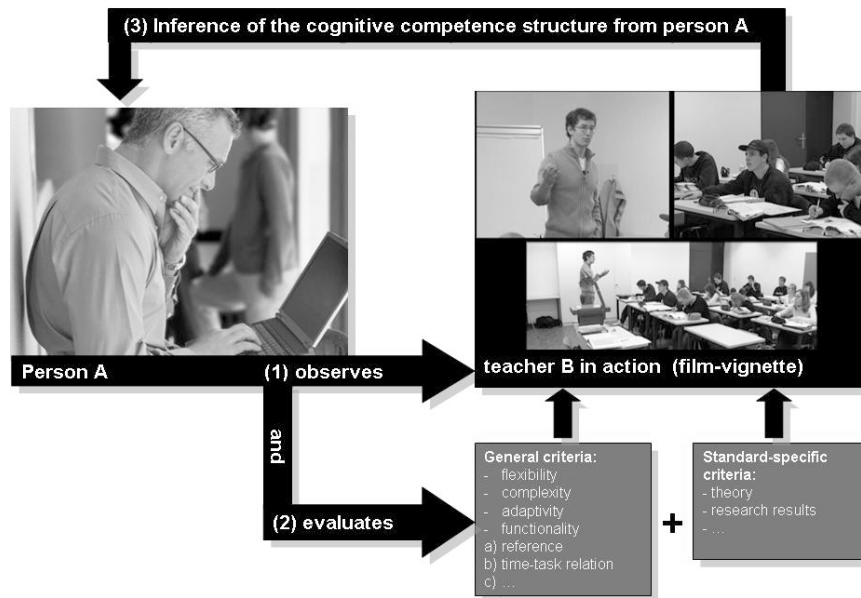
Figure 1: Observation and evaluation process for measuring competency profiles

experts, non-teachers) in the dimensions they judge. Because an important aspect of quality measurement consists of differentiation between evaluator groups, our diagnostic instrument predicts that teachers and non-teachers will differ significantly in their evaluation of the observed teaching behaviour in (almost) all dimensions. Our first modest expectation is that these differences are key indicators for competence, not performance.

### 1.3 On giving feedback: An omnipresent, but often unnoticed teaching standard

As an important example the competency profile (teaching standard) focused on in this paper is "The teacher is able to give supportive feedback to the students at different moments in time and by means of different tools". Oser et al. (2001) found that about 21% of the beginning teachers had never heard of this standard, 22% only in university courses, 41% in practical work and exercises, 12% had occasion to connect exercises related to this standard with theoretical reflection, and only about 4% had built a portfolio that employed this standard in applying theory to practice and in analytical exercises. Nearly all (94%) of these same students thought this standard was "very important" or "important", and the same percentage (95%) thought they would apply their knowledge with respect to this standard. This is despite the fact

that every teacher gives feedback dozens of times every day.

Feedback processes are omnipresent in instructional settings, whether intentional and goal-oriented or unintentional, verbal or non-verbal, oral or written. In this paper, we focus primarily on intentional verbal feedback from teachers. It is particularly characterized by its formative character, for in contrast to summative feedback - given at the end of a learning unit - verbal feedback *continuously* informs the students about the adequacy of their performance. Positive feedback makes learning success transparent, whereas negative feedback indicates a discrepancy between a certain learning goal and actual performance (Semmer & Pfäfflin, 1978; according to Volpert, 1973). If students do not know whether their efforts were successful, or if not, how they can improve, then learning is hindered or even rendered impossible.

Aside from this function as providing information, feedback also has a motivational impact. Supportive feedback motivates students to pursue current and future learning goals, and giving supportive feedback is of prime importance for motivating them to learn. Hence, both types of feedback - positive and negative - can be motivating, just as both types of information - corrective and confirming - are required in order to learn. However, if feedback is given in an offending, compromising way, it can have very negative effects on learning and motivation. If a teacher gives a student the impression he or she is thought unable to solve a given problem, success may become unlikely and motivation decrease. This could happen particularly if a student attributes the communicated inability as being a permanent state (attribution to ability).

Obviously, the difficulty and challenge of this teaching standard consists primarily in how to give *negative* feedback, in other words how to communicate when students make mistakes, when they fail, or when they are discouraged or hurt. But, when not perceived as a threat to self-esteem and thus ignored or rejected (Ilgen, Fisher & Taylor, 1979), negative feedback has a large potential to activate and support learning processes. Furthermore, it can motivate students to correct mistakes and to seek better approaches. This shows that the quality of feedback is crucial and that this teaching standard must be developed systematically in the education and further education of teachers. Thus, when we speak of "giving supportive feedback" as a teaching standard, we must define what the important criteria for supportive teacher feedback are. These criteria, together with other general dimensions of quality (referred to below as "cross-standard") in vocational education create a content-specific framework of a new, feedback-oriented, diagnostic instrument.

## 2.    Method

### 2.1    *Dimensions of the diagnostic instrument: standard-specific and cross-standard dimensions*

In the "advocatory approach", the diagnostic instrument (film vignette with subsequent questionnaire) is used to measure the sensitivity of a professional teacher.

There are two groups of dimensions that are considered as important here. One, which we call "standard-specific," asks about the aspects in each competency profile that are relevant to quality. As one can see in Table 1, and relevant to the particular professional standard we investigated, this can include such things as information content and precision ('specificity'), giving feedback regularly and frequently ('regularity'), explanation of reasons ('attribution'), and so forth.

Table 1: Standard-specific dimensions for giving supportive feedback

| Dimension | Content |
| --- | --- |
| Specificity | Information content and precision |
| Learning environment | Fearless instruction without abasements |
| Immediacy | Time difference between an answer or a performance of a student and the verbal feedback. |
| Regularity | Giving Feedback regularly and frequently |
| Practicability | Realistic goals and viable ways to achieve these |
| Attribution | Explanation of the reasons for a result and support of the students by using effort-oriented explanations |

The other dimension, which we call "cross-standard," includes more general criteria for the quality of vocational education that are meaningful for instructional processes as a whole. One example is the ability of a teacher to cope with the unforeseen or to manage spontaneous situations ('flexibility'), another is the teacher's sensitivity to the students ('empathy'), and so forth. We can capture the complexity of instruction more comprehensively only if we go beyond a view specific to a particular standard and include these "cross-standard" dimensions (see Table 2). These are derived in part from work done on defining aspects of high quality teaching (Einsiedler, 1997; Haenisch, 2000; Ditton, 2000; Helmke, 2003) and are intended to be valid for as many teaching standards as possible.

The "standard-specific" quality criteria of the 'feedback' competency profile are investigated using six dimensions, with two of them ("specificity" and "learning environment") - containing four sub-dimensions (see Table 4, stated below). These dimensions are largely derived from learning and motivation theories (Oser & Spychiger, 2005; Heckhausen & Rheinberg, 1980; Weiner, 1972), and are briefly described in the following.

Two aspects are emphasized by the term "*specificity*," the content of the information the teacher presents and the precision of the feedback given. We assume feedback is most effective when it is informative and if the information is focused on particular aspects of student performance relative to a given learning goal and the paths to reach it. The "*learning environment*" refers to aspects such as positive expectations, encouragement and appreciation of the teacher. The third dimension, "*immediacy*" refers to the time interval between student behaviour and teacher feedback. Because students need to link their performance to a corresponding feedback

Table 2: Cross-standard dimensions

| Dimension | Content |
|---|---|
| Empathy | Ability to empathize and the sensitivity to the students/trainees and their well-being |
| Effectiveness, division of time constraint | Aspects of efficiency<br>Comprehensibility, correctness and structure of the teaching |
| Commitment, motivation | Quality of motivation (teacher)<br>activity and commitment of students / trainees |
| Difficulty of tasks, adequacy of tasks | Appropriate degree of difficulty for requirements<br>Adequacy of content |
| Flexibility | Flexibility of teacher; coping with the unforeseen; spontaneous situation management |
| Autonomy, control | Appropriate control (by the teacher)<br>Support for self-reliance |
| Differentiation, complexity | Individual support of the teacher<br>content differentiation |
| Vision, importance of tasks | Aspects of teacher commitment (ethos: truthful/fair/considerate); Responsibility and role awareness |

from the teacher, we assume immediate feedback is more effective than delayed feedback. A delay risks fading memories about performance, which means the feedback information cannot be processed adequately (Semmer & Pfäfflin, 1978; London, 2003). The fourth dimension of "*regularity*" concerns the rate at which teacher feedback occurs. Receiving regular teacher feedback about the adequacy of performance is an important precondition for students to improve. No feedback and irregular feedback seem to hinder the learning and motivation processes considerably (London, 2003; Ilgen, Fisher & Taylor, 1979). "*Practicability*", the fifth dimension, means learning goals and the ways to reach them - on which teacher feedback focuses - must be adequate in terms of the required effort and the available time. The final dimension, "*attribution*", concerns causal inferences made by teachers in giving feedback. Such processes have an important impact on what learners believe (or come to believe) about themselves, or in other words, are about the inner attribution patterns of learners.

With respect to those six dimensions, the film and the questionnaire, we formulated the following hypotheses:

1. The theoretically developed dimensions (both "standard-specific" and "cross-standard") can be verified empirically

2. The "advocatory" approach will elicit differences between evaluator groups. Novice VET teachers and experienced VET teachers, we hypothesize, will evaluate the dimensions of the "feedback" diagnostic instrument differently

than will professional 1, professional 2 and non-teachers. We think this is probably due to different cognitive professional structures, though if differences between the evaluator groups can be found, this would be an important confirmation of the argument that "giving supportive feedback" is a profession-specific competency profile. By this argument, only teachers, and not laypersons or persons from completely different professions, know what aspects they need to consider in order to give supportive verbal feedback to young apprentices in an educational setting. Furthermore, differences in the evaluation of teachers and non-teachers are an important indicator of the specific professional knowledge of vocational school teachers.

### 2.2 Sample and procedure

We needed two samples to validate the 'feedback' diagnostic instrument, one of vocational school teachers and another of non-teachers. As for the vocational school teachers themselves, we contacted several vocational schools in the German-speaking part of Switzerland. The non-teachers were mainly recruited via associations such as choral societies or sports clubs. The sample therefore was not randomly chosen, though within the sample, different teaching subjects are represented and the competency profile was validated at four different vocational schools.

For both samples, data collection took place in the computer room of a school, each time with at least one researcher in attendance. The procedure is relatively lengthy, as participants needed approximately one hour to view the film vignette and answer all the questionnaire items. However, there were clear advantages: a high rate of return, the absence of missing data and the support that could be provided to the participants.

Participants watched the film vignette and rated the observed teaching behaviour directly on the computer. During the process of answering, participants could replay the film clip and could examine the documents used by the teacher in the class shown.

The sample of 107 persons were categorised (following the expertise models of Dreyfus & Dreyfus (1986) and Berliner (1992)) into five groups: non-teachers, advanced beginners (novice) with at most 3 years' teaching experience, experienced teachers 4-8 years' teaching experience, professionals 1 who had at least 9 years' worth of teaching experience, and professionals 2 with both at least 9 years' teaching experience and a diploma in practical mentoring (see Table 3 for more specific information about the sample).

All participants viewed the film clip "giving supportive feedback" and rated the observed teaching quality on a six-level scale (1 = *does not apply at all*; 2 = *does not apply*; 3 = *tends not to apply*; 4 = *tends to apply*; 5 = *applies*; 6 = *applies completely*) with respect to 'standard-specific' and 'cross-standard' dimensions.

Table 3: Distribution of the sample with some additional information

|  | advanced beginners | experienced teachers | profes- sionals 1 | profes- sionals 2 | non- teachers | total |
|---|---|---|---|---|---|---|
| **number** | 15 | 24 | 35 | 6 | 27 | 107 |
| **gender** |  |  |  |  |  |  |
| female | 5 | 4 | 5 | 1 | 19 | 34 |
| male | 10 | 20 | 30 | 5 | 8 | 73 |
| **age** |  |  |  |  |  |  |
| mean | 38.2 | 38.9 | 49.1 | 51 | 43.1 | 44 |
| (std.dev.) | (9.65) | (7.93) | (6.99) | (7.29) | (12.1) | (10.09) |
| min. | 22 | 27 | 33 | 42 | 18 | 18 |
| max. | 59 | 53 | 61 | 60 | 61 | 61 |

## 2.3 Statistical methods

Various statistical methods were used for data analysis. Hypothesis 1 required a confirmatory factor analysis followed by reliability estimation (Cronbach's alpha; split-half reliability, corrected according to Spearman-Brown). A t-test for independent samples as well as a variance analysis was used to test the second hypothesis. In a secondary analysis, the goodness of fit (in assigning individuals to the respective groups) was also undertaken, using a discriminant function.

## 2.4 The film vignette

The content of the film vignette consists of a five-minute-long setting in which a VET teacher, in a class of 17-year-olds in an agricultural school, asks these students to present their results after they have examined different types of motors. The students stand around the respective motor, and one student gives an oral and practical presentation of how one of the machines functions. The teacher always reacts to the presentation the same way. First, he gives positive feedback, then corrects the student and also provides new information.

## 3.  Results

The results of a confirmatory factor analysis (CFA) with the data from the aforementioned 107 participants showed that all the "standard-specific" and "cross-standard" dimensions can be found empirically, and that reliable scales can be generated.

Table 4: Results of the confirmatory factor analysis (CFA) (standard-specific dimension)

| Dimension | Number of items | Reliability* | Factor loading | Common variance(%) |
|---|---|---|---|---|
| Specificity | | | | |
| a) general specificity | 5 | 0.868 - 0.879 | 0.686 - 0.860 | 57.25% |
| b) current state | 4 | 0.716 - 0.749 | 0.503 - 0.822 | 44.50% |
| c) target state | 5 | 0.866 - 0.889 | 0.635 - 0.839 | 57.06% |
| d) bridging between current and target state | 5 | 0.794 - 0.833 | 0.521 - 0.824 | 51.32% |
| Learning environment | | | | |
| a) general | 4 | 0.862 - 0.887 | 0.746 - 0.869 | 66.65% |
| b) positive expectations | 3 | 0.824 - 0.850 | 0.732 - 0.879 | 68.41% |
| c) encouragement | 5 | 0.879 - 0.921 | 0.673 - 0.851 | 59.98% |
| d) appreciation | 5 | 0.872 - 0.903 | 0.574 - 0.863 | 59.75% |
| Immediacy | 3 | 0.850 - 0.861 | 0.678 - 0.942 | 67.52% |
| Regularity | 4 | 0.832 - 0.866 | 0.621 - 0.967 | 65.10% |
| Practicability | 5 | 0.852 - 0.855 | 0.715 - 0.751 | 53.76% |
| Attribution | 4 | 0.731 - 0.761 | 0.627 - 0.730 | 44.46% |

Note: *Cronbach's alpha and split-half, corrected in accordance with Spearman-Brown

Table 5: Results of the confirmatory factor analysis (CFA) (cross-standard dimension)

| Dimension | Number of items | Reliability* | Factor loading | Common variance (%) |
|---|---|---|---|---|
| Empathy | 8 | 0.915 - 0.931 | 0.530 - 0.908 | 58.64% |
| Effectiveness, division of time constraint | 3 | 0.547 - 0.631 | 0.441 - 0.738 | 39.14% |
| Commitment, motivation | | | | |
| a) teacher | 3 | 0.810 - 0.845 | 0.715 - 0.952 | 66.03% |
| b) students | 4 | 0.817 - 0.841 | 0.656 - 0.860 | 53.84% |
| Difficulty of tasks, adequacy of tasks | 4 | 0.619 - 0.727 | 0.433 - 0.774 | 43.16% |
| Flexibility | 6 | 0.902 - 0.914 | 0.641 - 0.894 | 64.59% |
| Autonomy, control | 5 | 0.804 - 0.828 | 0.593 - 0.827 | 50.81% |
| Differentiation, complexity | 4 | 0.851 - 0.853 | 0.653 - 0.846 | 59.58% |
| Vision, importance of tasks | 8 | 0.751 - 0.838 | 0.494 - 0.800 | 41.38% |

Note: *Cronbach's alpha and split-half, corrected in accordance with Spearman-Brown

Cronbach's alpha coefficients range from 0.716 to 0.921 for the "standard-specific" (Table 4) and from 0.631 to 0.931 for the "cross-standard" dimensions (Table 5).

A t-test for independent groups shows that the 'feedback' diagnostic instrument differentiates between vocational school teachers and non-teachers on eight of 12 "standard-specific" and on eight of nine "cross-standard" scales ($p < 0.05$) (see Tables 6 and 7). Highly significant differences between these two groups can be found ($p \leq 0.01$) on four of the "standard-specific" and six of the "cross-standard" scales and subscales. On all dimensions, vocational school teachers tend to evaluate the observed teaching behaviour more critically than do non-teachers, which means they show lower values on six-level scale. Furthermore, discriminant analysis shows that the correct allocation of teachers and non-teachers to the respective group succeeds in 91.6% of the cases. The canonical discriminant function is highly significant ($p \leq 0.01$).

Table 6: Results of the t-test for independent groups (standard-specific dimension)

|  |  | Mean | Std. deviation | t-value |
|---|---|---|---|---|
| **Specificity** | | | | |
| a) general | N-LP | 4.0444 | 0.88289 | |
| | LP | 3.6175 | 0.99528 | 1.98** |
| b) current state | N-LP | 4.0278 | 0.71500 | |
| | LP | 3.7781 | 0.94131 | 1.25 |
| c) target state | N-LP | 4.1556 | 0.67785 | |
| | LP | 3.5875 | 0.97271 | 3.34*** |
| d) bridging between current and | N-LP | 3.7852 | 0.85382 | |
| target status | LP | 2.7550 | 0.78851 | 5.74*** |
| **Immediacy** | N-LP | 4.1852 | 1.00992 | |
| | LP | 2.8833 | 1.13517 | 5.29*** |
| **Regularity** | N-LP | 4.1574 | 0.72771 | |
| | LP | 3.4688 | 1.03451 | 3.79*** |
| **Practicability** | N-LP | 3.8889 | 0.79097 | |
| | LP | 3.1750 | 0.85113 | 3.83*** |
| **Learning environment** | | | | |
| a) general | N-LP | 4.3056 | 0.84448 | |
| | LP | 3.9781 | 0.96472 | 1.57 |
| b) positive expectations | N-LP | 4.6790 | 0.61659 | |
| | LP | 4.3292 | 0.95097 | 2.19** |
| c) encouragement | N-LP | 4.6000 | 0.64926 | |
| | LP | 3.8850 | 0.93904 | 4.38*** |
| d) appreciation | N-LP | 4.5407 | 0.75613 | |
| | LP | 4.2900 | 0.93262 | 1.26 |
| **Attribution** | N-LP | 3.9352 | 0.87594 | |
| | LP | 3.6500 | 0.87385 | 1.46 |

Note: *$p \leq 0.1$, **$p \leq 0.05$; ***$p \leq 0.01$

Table 7: Results of the t-test for independent groups (cross-standard dimension)

|  |  | Mean | Std Deviation | t-value |
|---|---|---|---|---|
| Empathy | N-LP | 4.4907 | 0.85761 |  |
|  | LP | 3.9250 | 0.87004 | 2.93*** |
| Effectiveness, division of time constraint | N-LP | 4.1358 | 0.76939 |  |
|  | LP | 3.7375 | 0.77167 | 2.32** |
| Commitment, motivation (teacher) | N-LP | 4.6543 | 0.98485 |  |
|  | LP | 3.7500 | 1.03184 | 3.98*** |
| Commitment, motivation (students) | N-LP | 3.8519 | 0.91793 |  |
|  | LP | 3.3750 | 0.79556 | 2.58** |
| Difficulty of tasks, adequacy of tasks | N-LP | 4.2130 | 0.71288 |  |
|  | LP | 4.1188 | 0.82579 | 0.53 |
| Flexibility | N-LP | 4.3889 | 0.87217 |  |
|  | LP | 3.8979 | 0.94634 | 2.37** |
| Autonomy, control | N-LP | 4.1259 | 0.76891 |  |
|  | LP | 3.6275 | 0.92872 | 2.51** |
| Differentiation, complexity | N-LP | 4.0648 | 0.76458 |  |
|  | LP | 3.2813 | 0.89688 | 4.06*** |
| Vision, importance of tasks | N-LP | 4.7269 | 0.62600 |  |
|  | LP | 4.3063 | 0.73344 | 2.66*** |

Note: $*p \leq 0.1$, $**p \leq 0.05$; $***p \leq 0.01$

Unlike the significant differences between non-teachers and teachers in general, no significant differences could be found between the judgements of groups based on different degrees of teaching experience. The assumption that more experienced teachers differentiate between quality dimensions to a greater extent than do beginners is refuted, at least for this sample.

## 4. Reference frameworks for the measurement and the development of teaching competencies

At the beginning of this article, we spoke about three forms of measurement references with respect to quality. Here we only used a comparison of dependent variable means. The comparison of a single teacher's evaluation with the mean competency profile of a representative sample of vocational school teachers allows for an assessment vis-à-vis the professional field. Following this process, this teacher examined knows immediately if his or her own competency profile lies above or below the mean.

Still, this does not fully answer the question of quality, because the means calculated on different dimensions of this sample of vocational school teachers cannot be the quality norm. They are only reference points that help with respect to a given competency profile. Quality signifies something other than just an average character, since it has a normative aspect pointing toward a desirable direction and output.

The intent in calculating the mean is to analyse the fit between self-evaluation of observed teaching behaviour and an average frame of reference. We believe that such sensitisation, if not conceptual change, can be initiated in this manner, and thus professional behaviour optimized.

In order to establish the "real" quality of a given competency profile, a further frame of reference is needed. This would be an expert profile, a quality scale set through an expert rating. This assessment is negotiated among experts using a discursive (if not recursive) procedure and is the result of a consensus-finding process. What is more important than an examination of quantitative differences or similarities between the values of experts, teachers and non-teachers[2] is a visualisation or reflection on the justificatory structure of an expert's judgment.

In this way, prospective vocational school teachers can not only compare themselves with the mean of our normed sample of teachers, but can also come to recognise how experts evaluate teaching behaviour they observe and what reasons they give for their assessments. These considerations are important because experts refer to theories and base their judgments in part on meta-level considerations. Thus, existing perceptions are extended and new views as well as a deeper understanding are made possible. In combination, and with reference to one another, the reference frameworks provided by the 'feedback' diagnostic instrument open some innovative options for self-diagnosis. Together with attendant educational units, which link theory to exercise and practice, we can begin with the professional development of competency.

As to the competency profile "giving supportive verbal feedback", the expert rating took place with three persons, two renowned professors from German-speaking universities and a psychologist with a diploma in feedback. In a first step, they evaluated each item themselves. In the next stage, one expert read out an item, delivered its evaluation, and gave an explanatory statement to provide a basis for an ensuing discourse and consensus-building process (e.g., overcoming subjectivity). For the experts, the possibility was open that they might not reach consensus. The expert judgments at the level of items were subsequently summarized at the scale level. By means of this procedure, a quality scale for the measurement of competency profiles could be set.

The results of the expert rating permit initial comparative considerations, though these are neither conclusive nor based on statistically significant mean differences and must be interpreted descriptively. In Figure 2, no uniform pattern can be discerned when comparing the mean competency profile of non-teachers, vocational school teachers and the experts-consensus with one another on the "standard-specific" scales.

---

[2]Due to different methods of collecting data, a mean comparison can not be supported statistically either.
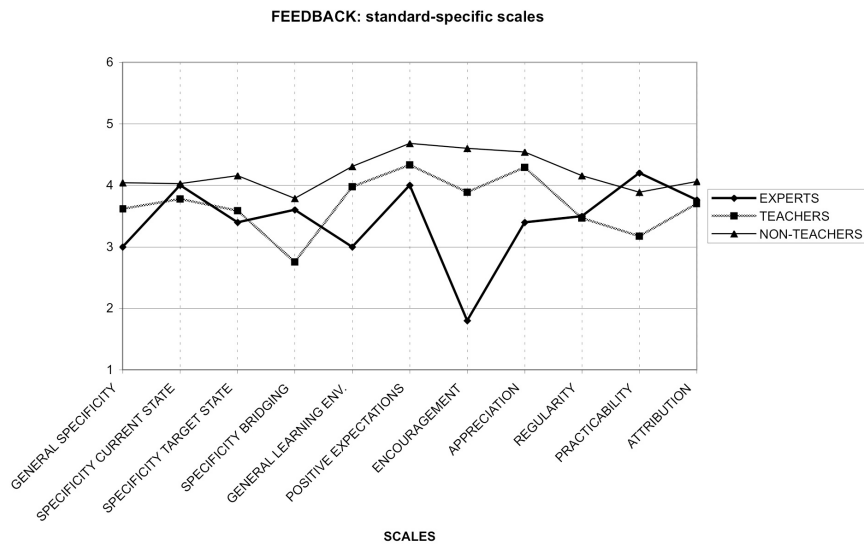
Figure 2: Expert rating: standard-specific scales

The results show that on six of 12 standard-specific scales and sub-scales[3], the experts valued the observed behaviour at the lowest level, as compared to the judgement of teachers and non-teachers. By comparison with the non-teachers experts evaluate the observed teaching behaviour more critically on all standard-specific scales and sub-scales, except for the scale "practicability" (adequacy of feedback about learning goals and achieving them). On this scale, the experts' evaluation lies slightly over the mean of the non-teachers as well as the teachers. It is the only dimension on which the experts show the highest value among the three groups. In general, the judgements of non-teachers show the most uniform pattern across all standard-specific scales and sub-scales whereas the teachers answered more differentiating. The experts show a peculiar zigzag-pattern. As for the cross-standard scales and sub-scales, evaluations are distinct by group (see Figure 3). The experts' judgment is markedly jagged.

On all the cross-standard scales, non-teachers' evaluation values are the highest, teachers' evaluation values lie under it, and the experts evaluated the observed teaching behaviour most critically. The experts therefore rated the teaching behaviour they observed, following criteria of general teaching quality, negatively. The differences

---

[3]The standard-specific scale "immediacy" was dropped as a result of a lack of consensus over one item of this scale.
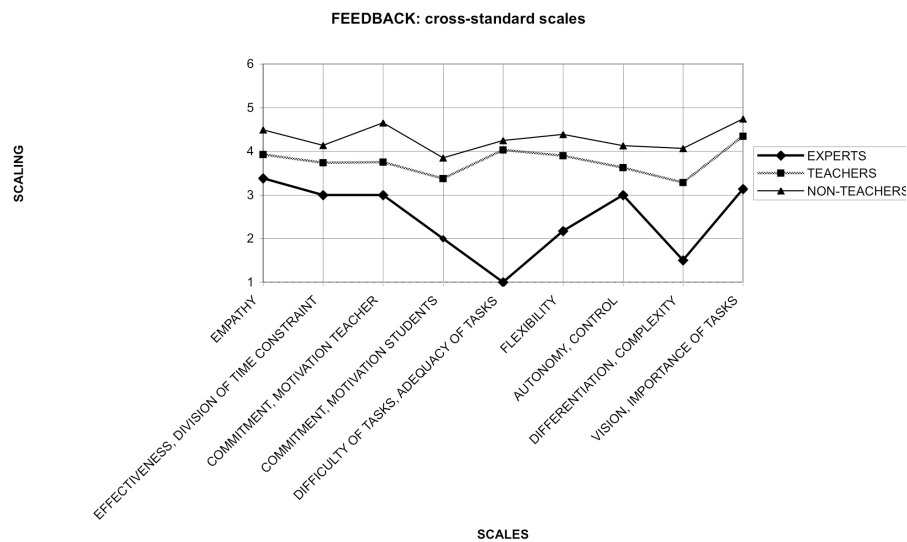
Figure 3: Expert rating: cross-standard scales

between the expert's standards and the mean competency standard of the teachers tend to be larger than the differences between the teachers' and the non-teachers' mean competency standard. However, the dimensions are often not directly observable, and the results indicate that experts are not only able to exactly differentiate between the specific dimensions of "giving supportive feedback" but also between the constitutive criteria of general teaching quality. Generally, it can be concluded that the experts tend to judge in a more differentiated manner and with greater sensitivity - but also with far greater negativity in their judgment.

## 5.    Discussion

As the results of the confirmatory factor analysis show, there is empirical support for the framework of the 'feedback' diagnostic instrument. Work on the identified dimensions can be continued, though the "standard-specific" dimension of '*immediacy*' was dropped due to a lack of expert consensus on one item of this scale. This does <u>not</u> mean immediacy is an unimportant dimension in giving supportive teacher feedback, but rather that the scale and the corresponding items must be improved and be better adapted to the film vignette. Teachers and non-teachers differ significantly in their evaluation of the scales and sub-scales of the 'feedback' diagnostic instru-

ment, which indicates that "giving supportive feedback" is a specific professional ability within a school setting.

However, such differences were not evident between the various teacher groups who evaluated. Possible reasons include that the sample was too small to find significant differences, that the criterion for categorisation (number of years' teaching experience) is not as critical for competence as supposed, or that there were too many teachers without a teaching diploma (approx. 50) in the sample. Other reasons might be explain this non-finding as well.

Still, despite what is an only partial confirmation of the second hypothesis, the differences seen between teachers and non-teachers provides important confirmation of professional knowledge as well as an important empirical basis for a competence-oriented professionalization of the education of vocational school teachers.

For this to be the case, quality has to be defined for the competency profile, and it was for that reason that an expert rating was conducted in order to establish a quality scale. This is one possible avenue, and suggests an interesting use for future teachers' education, though none conclusive explanations can be given so far for the discrepancies between the judgment by experts and judgments by other groups. One can assume experts have highly specific knowledge, and that this is what leads their evaluation to be more differentiated as well as more sensitive to the observed situation. In qualitative statements, the experts expressed their understanding for the situation and admitted that on some "standard-specific" scales, the observed teacher showed behaviour adequate to the situation. Teachers and non-teachers, however, estimated the different scales in a uniform and not very diverse pattern. By visualizing the distance between teacher, non-teacher and expert judgment, or by including qualitative expert statements, the development of teacher competence could be enhanced. Of course, this assumption has first to be tested empirically.

## 6. Conclusions

The 'feedback' diagnostic instrument with its quality-related frames of reference is a prototype for the self-diagnosis of the quality of a defined competency profile, which - together with attendant educational units - can trigger the development of professional competence. The consultation with experts, as a third frame of reference, represents a quality norm, inasmuch as it takes a prescriptive function in the process of professionalization - which may or may not be desirable. Of course, the procedure for measuring vocational schoolteachers' competence, as exemplified here, is at present only in an initial phase. Further efforts must be undertaken to improve this prototype and to answer questions about quality scales for measuring competencies in a more differentiated way. Thus, further expert ratings may be used to examine problems associated with the consensus-building process (such as the gradual adjustment of assessments that lead to conformity). Furthermore, perhaps one can identify specific clusters within the group of teachers or identify individual cases that have remarkable competency profiles - this is a different meaning of 'quality' -

could be used in the training of vocational school teachers. In addition, diagnostic instruments for additional competency profiles have to be developed and validated, as well as implemented in the field. Our diagnostic instrument will be re-used for a pilot intervention study in the education of vocational school teachers. It is particularly interesting that not only will the development of competencies be measured, but so will the effects on teachers acting in real classroom situations. Thus, we will be breaking new ground in vocational training research.

## References

Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9 (4), 469-520.

Beck, K. (2006). Standards - ein Mittel zur Qualitätsentwicklung in der Lehrerbildung?. *Wirtschaft und Erziehung*, 56, 44-54.

Beck, K. & Dubs, R. (1998). Kompetenzentwicklung in der Berufserziehung: Kognitive, motivationale und moralische Dimensionen kaufmännischer Qualifizierungsprozesse. In K. Beck & R. Dubs (Eds.) *Zeitschrift für Berufs- und Wirtschaftspädagogik (ZBW)*, 14. Stuttgart: Steiner.

Berliner, D.C. (1992). The nature of expertise in teaching. In F. Oser, A. Dick & J.-L. Patry, (Eds.) *Effective and Responsible Teaching: The New Synthesis*. San Francisco: Jossey-Bass.

Borich, G.D. (1977). *The appraisal of teaching: concepts and process*. Reading, MA: Addison-Wesley.

Bromme, R. (1992). *Der Lehrer als Experte. Zur Psychologie des professionellen Wissens*. Bern: Hans Huber.

Darling-Hammond, L.; Hammerness, K.; Grossmann, P.; Rust, R. & Shulman L. (2005). The design of teacher education programs. In L. Darling-Hammond & J. Bransford (Eds.) *Preparing Teachers for a Changing World*. San Francisco: Jossey-Bass.

Ditton. H. (2000). Qualitätskontrolle und -sicherung in Schule und Unterricht. Ein Überblick über den Stand der empirischen Forschung. *Zeitschrift für Pädagogik*. (41. Beiheft), 73-92.

Dreyfus, H.L. & Dreyfus S.E. (1986). *Mind over machine. The power of human intuition and expertise in the era of the computer*. New York: The Free Press.

Einsiedler, W. (1997). Unterrichtsqualität und Leistungsentwicklung. Literaturüberblick. In Weinert, F. E. & Helmke, A. (Eds.) *Entwicklung im Grundschulalter*. Weinheim: Psychologie Verlags Union.

Fahrenberg, J., Myrtek, M., Pawlik, K. & Perrez, M. (2007). Ambulantes Assessment - Verhalten im Alltagskontext erfassen. Eine verhaltenswissenschaftliche Herausforderung an die Psychologie. *Psychologische Rundschau*, 58 (1), 12-23.

Frey, A. (2006). Methoden und Instrumente zur Diagnose beruflicher Kompetenzen von Lehrkräften - eine erste Standortbestimmung zu bereits publizierten Instrumenten. *Zeitschrift für Pädagogik*, (51. Beiheft), 30-46.

Frey, A. & Balzer, L. (2005). Der Beurteilungsbogen smk. In A. Frey, R.S. Jäger & U. Renold (Eds.) *Kompetenzdiagnostik. Theorien und Methoden zur Erfassung und Beurteilung von beruflichen Kompetenzen*. Landau: VEP.

Haenisch, H. (2000). Merkmale erfolgreichen Unterrichts. Forschungsbefunde als Grundlage für die Weiterentwicklung von Unterrichtsqualität. In Landesinstitut für Schule und Weiterbildung (Eds.) *Was ist guter Unterricht?*. Bönen: Druckverlag Kettler.

Hammerness, K., Darling-Hammond, L., Grossmann, P., Rust, F. & Shulman L. (2005). The design of teacher education programs. In L. Darling-Hammond & J. Bransford (Eds.) *Preparing Teachers for a Changing World*. San Francisco: Jossey-Bass.

Heckhausen, H. & Rheinberg, F. (1980). Lernmotivation im Unterricht, erneut betrachtet. *Unterrichtswissenschaft*, 1, 7-47.

Helmke, A. (2003). *Unterrichtsqualität - erfassen, bewerten, verbessern*. Seelze: Kallmeyersche Verlagsbuchhandlung.

Helmke, A. & Schrader, F.-W. (1993). Was macht erfolgreichen Unterricht aus? Ergebnisse der Münchner Studie. *Praxis Schule*, 5-10 (1), 11-13.

Ilgen, D. R.; Fisher, C. D. & Taylor, S. M. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology*, 64 (4), 349-371.

Kounin, J. (1976). *Techniken der Klassenführung*. Bern: Huber Klett.

London, M. (2003). *Job feedback. Giving, seeking, and using feedback for performance improvement. (2nd ed.)*. Mahwah, N.J.: L. Erlbaum.

Max, Ch. (1999). *Entwicklung von Kompetenz - ein neues Paradigma für das Lernen in Schule und Arbeitswelt*. Frankfurt am Main: Peter Lang GmbH.

Neuenschwander, M. (2004). Lehrerkompetenzen und ihre Beurteilung. *Journal für lehrerinnen- und lehrerbildung*, 1, 23-29.

Nickolaus, R. (2007). Qualität in der beruflichen Bildung. *Zeitschrift für Berufs- und Wirtschaftspädagogik (ZBW)*, 103 (2), 161 - 166.

Oser, F. (2001). Standards. Kompetenzen von Lehrpersonen. In F. Oser & J. Oelkers (Eds.) *Die Wirksamkeit der Lehrerbildungssysteme. Von der Allrounderbildung zur Ausbildung professioneller Standards*. Chur; Zürich: Rüegger Verlag.

Oser, F. (2007). Willkür als Feind der Spontaneität. Aspekte der Standardisierung des Lehrerhandelns. In D. Benner (Eds.) *Bildungsstandards. Instrumente zur Qualitätssicherung im Bildungswesen. Chancen und Grenzen - Beispiele und Perspektiven*. Paderborn; München; Wien; Zürich: Ferdinand Schöningh.

Oser, F.; Curcio, G.-P. & Düggeli, A. (2007). Kompetenzmessung in der Lehrerbildung als Notwendigkeit - Fragen und Zugänge. *Beiträge zur Lehrerbildung*, 25 (1), 14-26.

Oser, F., Curcio G.-P., Düggeli, A. & Kern M. (2006). *Schlussbericht 1. Phase. Professional Minds - Handlungssituationen und Standards für die Berufsausbildenden (6376.1BFS)*. University of Fribourg unpublished working paper.

Oser, F. & Heinzer, S. (2008). Die Entwicklung eines Qualitätskonstrukts zur advokatorischen Erfassung der Professionalität. *mimeo*.

Oser, F. & Spychiger, M. (2005). *Lernen ist schmerzhaft. Zur Theorie des Negativen Wissens und zur Praxis der Fehlerkultur*. Weinheim; Basel: Beltz.

Seidel, T. & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: the role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77(4), 454-499.

Semmer, N. & Pfäfflin, M. (1978). *Interaktionstraining: Ein handlungstheoretischer Ansatz zum Training sozialer Fertigkeiten*. Weinheim; Basel: Beltz.

Shulman, L.S. (1986). Those who understand: knowledge growth in teaching. *Educational Leadership*, 15, 4-14.

Shulman, L.S. (2007). Signature Pedagogies in the Professions. In D. Lemmermöhle, M. Rothgangel, S. Bögeholz, M. Hasselhorn & R. Watermann (Eds.) *Professionell lehren, erfolgreich lernen*. Münster: Waxmann.

Volpert, W. (1973). *Sensumotorisches Lernen*. Frankfurt am Main: Limpert.

Weiner, B. (1972). *Theories of motivation*. Chicago: Markham.

Weinert, F.E. (2001). Concept of Competence: A Conceptual Clarification. In D. Rychen & L. H. Selganik (Eds.) *Defining and selecting key competencies*. Seattle, Toronto, Bern Göttingen.

Weinert, F. E. & Helmke, A. (1996). Der gute Lehrer: Person, Funktion oder Fiktion?. In A. Leschinsky (Eds.) *Die Institutionalisierung von Lehren und Lernen, Zeitschrift für Pädagogik, (34. Beiheft)*. Weinheim: Beltz.

Zeichner, K. (2006). Konzepte von Lehrerexpertise und Lehrerausbildung in den Vereinigten Staaten. *Zeitschrift für Pädagogik*, 2, 97-113.