

## Teacher Competence: assessment for Quality

Erik Roelofs  
Cito, Arnhem

Keynote-paper presented at the Conference of the Association of Educational Assessment-Europe, November 9<sup>th</sup> 2007, Stockholm

### Abstract

*The development of instruments to assess teacher competence requires a model of competent performance which can guide both the collection and appraisal of evidence in task situations. Following Kane (1992), the validation of statements about teachers' competence is regarded as the evaluation of interpretive argumentation.*

*On the basis of contemporary insights into teaching and learning, an interpretive model of competent performance is described which rather than being prescriptive in nature offers a scope for various forms of responsible professional performance. Consequences of professional performance for students/class/organization are the basis of the model. Acceptable interventions and underlying decision-making processes as well as the associated parts of a professional knowledge base are derived from the consequences. The consequences of these insights for the elaboration of domains of competence and the collection of evidence are discussed.*

### 1. Teacher quality

Internationally, a growing body of research shows that the quality of the teacher in the classroom is the most important schooling factor predicting student outcomes (e.g. Darling-Hammond, 2000, 2003; Santiago, 2002). Based on reviews of studies of student achievement in the United States, Darling-Hammond, LaFors, & Snyder (2001) concluded that teachers' qualifications based on measures of knowledge and expertise, education, and experience account for a larger share of the variance in students' achievement than any other single factor, including poverty, race, and parent education (p.10). It is reported that students who are assigned to several ineffective teachers in a row have significantly lower achievement and gains in achievement than those who are assigned to several highly effective teachers in sequence (Sanders & Rivers, 1996).

Given these empirical findings it is not surprising that throughout the world policymakers, educators and researchers are trying to capture the contents of these teachers' qualifications within professional standards. Since the enactment of the No Child Left Behind act and the Higher Education Act (HEA), the United States Department of Education promotes stronger standards and accountability in teacher preparation, teacher quality and teacher certification (U.S. Department of Education Office of Postsecondary Education, 2005). Similar developments have taken place in other countries. An important consequential activity is to develop competency frameworks of quality teaching that serve different purposes: first to provide a template for designing courses for pre-service training. Second, to provide the context for informed performance management discussions. This process involves supporting and empowering staff, helping staff plan and evaluate their work, and providing opportunities for growth and development. Third, related to the first purposes a competency framework can serve professional development. Teachers and line managers can negotiate to use it for determining and prioritising areas of professional growth, assisting in the selection of professional development activities, and informing ongoing personal and career development planning. Fourth for individual teachers a framework allows them to make judgements about the effectiveness of their teaching and to adjust their practice accordingly. Fifth, it a competency framework may serve as a basis for teacher assessment.

Articulating professional standards for teaching helps to make the knowledge and capabilities of teachers explicit for those within and outside the profession, and provides means by which good teaching can be identified, rewarded and celebrated (National Reference Group for Teacher Standards Quality and Professionalism, 2003; Ramsey, 2000). Moreover, it is increasingly agreed that identifying and publicly recognizing what it is that effective teachers know, do and value is an important step in enhancing the public profile and standing of the profession.

Several examples of frameworks for teacher competence have been developed recently. A well known competency framework for teachers was developed by the United States National Board for Professional Teaching Standards (2001). Based on five core propositions frameworks have been developed for teachers teaching different subjects and different age groups. The propositions entailed:

- 1) Teachers are committed to students and their learning.
- 2) Teachers know the subjects they teach and how to teach those subjects to students.
- 3) Teachers are responsible for managing and monitoring student learning.
- 4) Teachers think systematically about their practice and learn from experience.
- 5) Teachers are members of learning communities.

Similar frameworks have been developed in other countries, e.g. in Australia (Department of Education and Training, 2004) and in the Netherlands. The Dutch foundation for professional teaching competence (SBL, 2003) formulated requirements for seven domains of competence that are considered crucial for all teachers (see also section 3.2.2.1).

Frameworks for teacher competence, and more specifically standards for teacher competence, have formed the basis of various instruments to assess teachers at various stages in their professional careers in the context of selection, certification, and professional development.

However, there are many conceptions of what is referred to as teacher competence. The questions in this paper are, how do these conceptions look like and how can teaching competence be conceptualized to form a basis for valid, reliable, acceptable and affordable instruments?

To address these questions, some fundamentals of a framework for the assessment of teacher competence will be presented in this paper. We start by deriving an interpretive model for the assessment of competent (teacher) performance, based on different theoretical notions of good teaching. In the second part of the paper, we discuss various aspects of instruments for assessing teacher competence. This is done using the requirements that follow from the concept of construct validity as expounded by Messick (1996). Lastly, we discuss some issues for future study relating to the assessment of teacher competence.

## **2. An interpretive model of teacher competence**

There is no generally accepted definition of the concept of competence. Recently several Dutch authors (e.g., Bos, 1998; Mulder, 2001; Van Merriënboer, Van der Klink & Jansen, 2002) have reviewed the literature on this concept and came up with comprehensive definitions. A first and important distinction can be made between 'competence' and 'competency'. According to Mulder, competence is a comprehensive concept for abilities or capabilities of people or organizations, while a specific competency forms a part of competence. Competency (plural competencies) is a narrower, more atomistic concept used to label particular abilities (see also McConnell, 2001). On the basis of a study of dozens of definitions of competence (e.g., Bunk, 1994; Spencer & Spencer, 1993; Parry, 1996), Mulder (2001) derived a definition that captures most of the important authors: 'competence is the ability of a person or organization to achieve particular levels of performance' (p. 76). Citing different authors he adds that the competencies of individuals consist of:

- (a) integrated action proficiencies
- (b) which are made up of clusters of knowledge structures,
- (c) cognitive, interactive, emotional, and where necessary psychomotor skills,
- (d) and attitudes and values which are necessary for
- (e) the performance of tasks,
- (f) solving of problems,
- (g) and more generally the ability to function in a particular occupation,
- (h) a particular organization,
- (i) a particular position,
- (j) or a particular role.

When measuring dimensions of competence, it must be realized that these are not directly observable, but are manifested in performance in a specific situation (Spencer & Spencer, 1993). In addition, competence can be developed to a particular level, for example beginner, advanced, and expert. Mulder (2001) emphasizes that competence may be present in individuals (personal

competence) and systems (system or team competence). Finally, aspects of competence are to some extent transferable from one situation to another (Thijssen, 1998, 2001).

Various questions can be asked about the assessment of the competence of individuals. How are statements about competence derived? What assumptions and theoretical notions underlie measurements of competence? In answering questions of this type, it is important to use an adequately descriptive and explanatory interpretive model (Shepard, 1993).

There is no sound and broad-based scientific framework of what constitutes competent teaching from which inferences can be drawn for the purposes of assessing teacher competence (Haertel, 1991). There are various frameworks, the contents of which are largely dependent on the underlying vision of professional performance (Dwyer, 1994, 1998) and on the theoretical approaches adopted (cf. Reynolds, 1992).

Developers of teacher assessment instruments mostly work towards a shared view of competent teaching, obtained through interaction between developers and representatives of the profession. The resultant view of this interaction can vary widely: it may be a hybrid of all kinds of views of teaching, but also a fairly specific view, for example, 'program-oriented' or 'development-oriented learning' in the context of early and pre-school education. The first approach is suitable to adopt in the formulation of frameworks of competent performance which must apply to large groups of teachers, for example, national standards of teaching. The second approach is more appropriate for organizations that work according to a specific mission.

Besides the view of teaching, the theoretical angle on professional performance also determines what form an interpretive model takes. In the literature, different elements of teacher competence have been emphasized throughout the history of evaluating teachers. In reviewing the literature, different conceptions of good teachers and good teaching can be distinguished (Creemers, 1991; Verloop, 1999) :

- (a) the differentiation of personality traits which help to make a successful teacher (Getzels, & Jackson, 1963; Creemers (1991);
- (b) the description of knowledge elements involving subject matter content, ways teachers of thinking within a discipline (Bruner, 1963; Tom & Valli, 1990);
- (c) the description of forms of teacher behavior which contribute to learning performance (Brophy & Good, 1986; Simon & Boyer, 1974);
- (d) the description of the cognitions and decision-making processes of teachers (Kagan, 1990; Verloop, 1988);
- (e) the description of the practical knowledge of teachers which does justice to the specific practical situations in which individual teachers find themselves (their class, their subject domain) and the way in which they form theories about these situations (Beijaard & Verloop, 1996)

For each of these conceptions of good teaching, specific kinds of assessment techniques were used. To assess personality traits, questionnaires and psychological tests were used to identify certain desirable or undesirable traits. Within this conception of teaching the focus was not so much aimed at good teaching itself but rather at characteristics of a good citizen (Getzels, & Jackson, 1963).

A conception that is still dominant is that good teachers have a lot of knowledge. In earlier times, this involved knowledge of discrete facts and elements; later emphasis was laid on the structure of a discipline (like math or physics), and still later on how professionals act and think within a certain discipline. Knowledge also relates to pedagogical knowledge, for instance about methods of delivering instruction, building curriculum, and grouping students. These knowledge elements are increasingly derived from educational research (Bellon, Bellon & Blank, 1990). A frequently used method of assessing knowledge has long been and still is taking standardized knowledge tests (e.g., Latham, Gitomer & Ziomek, 1999). A well known example in the United States were the National Teacher Examinations (NTE; see Haertel, 1988). The tests measured communication skills, covering reading, writing, and listening; general knowledge, covering literature and fine arts, basic mathematics, science, and social studies; and professional knowledge, covering planning for instruction, implementing instruction, and evaluating learning outcomes, as well as laws affecting classroom practice, extra classroom influences on children, and knowledge about the teaching profession and standards of professional conduct.

As a reaction to both the emphasis on stable teacher characteristics and a one-sided emphasis on knowledge, teaching is also seen as displaying effective behavior. This conception takes into account what a teacher actually shows in the classroom. Numerous observational instruments have been developed that concentrate on (small) units of behavior, which are thought to be connected with successful learning outcomes (Stodolsky, 1990). Strong research support has linked student learning to variables such as teacher clarity, enthusiasm, task-oriented behavior, variability of lesson approaches, and student opportunity to learn criterion material. Teachers' abilities to structure material, ask higher order questions, use student ideas, and probe student comments have also been found to be important variables in what students learn (Rosenshine & Furst, 1973; Darling-Hammond, Wise, & Pease, 1983; Good & Brophy, 1986). Another general finding was that no single instructional strategy appeared to be unvaryingly successful; instead, teachers who are able to use a broad repertoire of approaches skillfully (e.g., direct and indirect instruction, experience-based and skill-based approaches, lecture and small group work) are typically most successful.

A multitude of instruments and systems have been used for observing teachers' classroom performance, from sporadic, informal, and unsystematic observations to elaborate systems requiring carefully trained evaluators (Medley, Coker, & Soar, 1984). The Florida Performance Measurement System (FPMS) was a carefully designed and thoroughly validated instrument (Peterson, Micceri, & Smith, 1985). The FPMS is a low-inference observational system including formative (diagnostic) instruments for each of six domains, and the Summative Observation Instrument (SOI). Development of the FPMS began with a review of process-product research on teaching, which yielded 31 "concepts of effective teaching behavior" (Peterson et al., 1985, p. 64). These were organized into the six domains of ((a) instruction: planning; (b) management of student conduct; (c) instructional organization and development; (d) presentation of subject matter; (e) communication: verbal and nonverbal; and (0) testing: student preparation, administration, and feedback. The concepts within these domains were then operationalized by 124 indicators of effective or ineffective teacher behavior.

However, the 'effective teaching behavior' approach pays little attention to what goes on in the mind of the teacher. What does he think or decide, and why does he decide the way he does? Different assessment instruments have been used to uncover teacher thinking, including thinking aloud protocols while solving a teaching problem and administering stimulated recall interviews. During these interviews, teachers look back at their videotaped performance and answer questions about what they were thinking at a particular moment.

The concept of teaching as displaying a rich base of practical knowledge involves assessment methods that concentrate on specific situations the teacher is confronted with. The instruments used do not differ principally from those used in revealing thought processes. However, the focus is more on the specific work context (e.g., the specific subject in a specific grade) in which teachers carry out their activities (Meijer, Verloop & Beijaard, 1999).

Using the above conceptions of 'good teaching in assessments will result into direct measures of some teacher attributes, be it characteristics, knowledge, behavior, or thinking.

Finally, a relatively new way of conceiving teaching is to draw in the consequences of teaching for students' leaning activities and learning results. Until recently theories of teaching took little account of the results of research on learning processes (e.g. see Duffy, Lowyck & Jonassen, 1993). Theories of learning and theories of teaching often originate and operate independently from one another. Increasingly, the two types of theories are being combined into one theory. This may be conceived of as a new conception of teaching: the facilitation of learning activities, studying the relationships between teacher activities and learner activities including cognitive, affective, motivational, and developmental factors of learning (Shuell, 1993, Vermunt & Verloop, 1999). In this conception, competent teaching assumes that teachers do not so much need to demonstrate a clear-cut repertoire of 'correct' behaviors but rather show that their interventions contribute to successful learning activities on the part of students (Simons, 1999; Vermunt & Verschaffel, 2000).

In addition, developments within the context of accountability led to the 'value added approach' of teacher evaluation (Wright, Horn, and Sanders, 1997; Sanders, Saxton, & Horn, S.P., 1997). In this notion, students' gains on standardized achievement tests are used to evaluate individual teachers. The authors claim that the annual academic growth of students can be separated into two parts: that which can be attributed to the student and that which can be attributed to the individual classroom, school or district. Without crediting or discrediting this approach the notion can be considered as part of a view of good teaching: taking care of students' growth.

Assessing different domains of teacher competence means that interpretive inferences are made about teachers (cf. Kane, 1992) based on scores on instruments. Using the separate conceptions of teaching as a basis for teacher assessment may limit the scope of interpretive arguments about teachers competence. Combining them into one comprehensive model of competent performance, the chances for connected inferences are better than when using reductionistic models which concentrate on separate parts of the teaching process. Roelofs and Sanders (2007) have recently developed an eclectic model of competent performance for the assessment of teacher competence which captures the aspects of competent performance described above.

The starting point in this model, represented in Figure 1, is that teacher competence is reflected in the consequences of teachers' actions, the most important being students' learning activities. Other examples of consequences are: a (smooth or disruptive) classroom climate, a feeling of well-being among students, good relationships with parents and colleagues. Starting from the consequences, the remaining elements of the model can be mapped backwardly. First, the component 'actions' refers to professional activities, e.g., delivering instruction, providing feedback to students, and creating a cooperative classroom atmosphere. Second, any teacher activity takes place within a specific context in which a teacher has to make many decisions, on a long term basis (planning ahead) or immediately within a classroom situation (cf. Doyle, 1983). For instance, teachers will have to plan their instruction and adapt it depending on differing circumstances (e.g., different student learning styles, different organizational conditions). Third, when making decisions and performing activities, teachers will have to draw from a professional knowledge base and from some personal characteristics.

insert Figure 1 about here

---

Evaluating different domains of teacher competence, for example, instruction and classroom management, means that interpretive inferences are made about teachers performance (cf. Kane, 1992). When combining different aspects of teaching into one comprehensive model of competent performance, the chances for valid inferences are better than when using reductionistic models which concentrate on separate parts of the teaching process.

### **3. Assessment of teacher competence**

Before describing the implications of our general model for collecting evidence of teacher competence, we briefly comment on the importance of construct validity as a unifying concept for determining assessment quality. Many specific quality requirements can be derived from this framework of construct validity.

#### **3.1 Criteria for construct validity of competence instruments**

The requirements of competence instruments vary depending on the purpose of the assessment. Because of the consequences for the candidate, stricter requirements are set for 'high-stake' instruments (selection, certification) than for instruments used for professional development (Pelkmans, 1998).

The most extensive framework for the determination of the quality of instruments has been developed by Messick (1996). Messick indicates that for each form of assessment it is necessary to take into account six aspects of construct validity:

- (a) content,
- (b) theory and process models,
- (c) structure,
- (d) generalizability,
- (e) external aspects,
- (f) consequences.

The 'content' aspect is concerned with the relevance and representativeness of the assessment. The question is: within what limits can conclusions be drawn on the basis of the assessment? The 'theory and process models' aspect is concerned with the question of the extent to which the selected tasks call for the relevant action on the part of a candidate and whether the influence of construct-irrelevant factors is minimized. The 'structural aspect' relates to the question of whether the performance criteria

correctly reflect the criteria that experts use and the accuracy and consistency with which performance is scored and assessed.

The aspect of generalizability relates to the extent to which the assessments concerned can be generalized to a universe of, for example, tasks and settings.

The external aspects of validity relate to the extent to which the measurement results converge with, and diverge from, other measurements and constructs. The aspect of 'consequences' or consequential validity, examines the extent to which the instrument has positive or negative effects and side effects on the student's learning and the teacher's teaching.

### **3.2 Collecting evidence of competence**

Using our comprehensive model of teacher competence as an interpretive framework for assessment, and taking into account criteria for construct validity, consequences for the construction of a content domain and procedures for collecting evidence of competence can be described. Of the Messick aspects three deserve closer attention: content, theory and process models, and generalizability.

#### **3.2.1 The development of a domain of competence**

Following Messick's 'content' aspect, the assessment content should be relevant and representative for the teaching profession. Content reviews are used to set the boundaries within which inferences about teachers' competence are made. Various complementary procedures are usually adopted in establishing a domain of competence: empirical analyses of how teachers function, consultation of excellent teachers, empirical research on variables which contribute to higher learning performance, and consultation of committees of practicing professionals (cf. Verloop, Beijaard & Van Driel, 1998). The mix of scientific and practical perspectives contributes to the acceptance and practical usability of instruments (Beijaard & Verloop, 1996; Duke & Stiggins, 1990; Uhlenbeck, 2002).

The basis of any assessment is an elaborated domain of competence containing an overview of aspects of competence, the situations in which they must be demonstrated, and the desired degree of mastery. Three questions need to be answered:

- (a) what is the crucial content of competence?
- (b) how are performance criteria defined?
- (c) in what way can levels of competence be assessed?

##### *3.2.1.1 Selection of content*

The demarcation of domains of competence necessitates selection of what is characteristic of adequate professional functioning and what is critical to functioning. Starting from our model of competence, the important questions are:

- What are teachers expected to demonstrate and in which task situations?
- What degree of difficulty of task situations must teachers be able to cope with?
- What student results ('consequences') can be expected from teacher activities?
- Through which actions and decision-making processes might teachers be able to contribute to students' results?

Various domain descriptions for teacher competence have been developed in the United States as well as in the Netherlands that can be considered. Danielson and McGreal (2000) distinguished four broad, relevant professional task areas: planning and preparation; instruction; classroom environment; professional responsibilities. Referring to the domain classroom environment, they state: [...] Such activities and tasks establish a comfortable and respectful classroom environment, which cultivates a culture for learning and creates a safe place for risk taking."(pp. 31). With respect to instruction, they write: "[...] Teachers who excel in domain 3 [instruction] create an atmosphere of excitement about the importance of learning and the significance of the content." (pp. 32). These statements illustrate a focus on the consequences of the actions, rather than on the actions themselves.

The description given by Danielson et al. was also the basis for a set of assessment instruments, Praxis III, as part of the so-called PRAXIS series, developed by Educational Testing Service, measuring, among other things, teacher's in-class practice in each of the four areas mentioned. In the PRAXIS-III assessment, classroom observations of teacher and student behavior are combined with pre- and post-observation interviews (Dwyer, 1998), the latter addressing the decision-making process of teachers.

The American National Board for Professional Teaching Standards reduces the multiplicity of tasks in the assessment for the certification of (advanced) teachers to dimensions of teaching expertise, such as improvisation, degree of challenge, passion for teaching and learning (Bond, Smith, Baker, & Hattie, 2000).

In the Netherlands, the Dutch Foundation for Professional Teaching Competence (SBL, 2003) developed a set of initial proficiency requirements for teachers in primary and secondary education for the Ministry of Education on the basis of seven broad domains of competence. In the description of the requirements, SBL starts with describing how classes and individual students of competent teachers function, illustrating a coherent approach to teaching. The following domains have been elaborated: 1. Interpersonal competence, referring to the ability to create a friendly, cooperative climate along with open communication. 2. Pedagogic competence, referring to the ability to create a psychologically safe learning environment for students, contributing to their well-being. 3 Subject matter and didactic competence, referring to the ability to guide students in acquiring the basics of school subjects and the way this knowledge can be used in daily life and working life. 4. Organizational competence, referring to the ability to create an orderly on-task climate in their classes. 5. Competence in cooperating with colleagues, referring to the ability to gear one's own work to the work of colleagues and to contribute to the school organization in general. 6. Competence in cooperating with the school environment, referring to the ability to contribute to the cooperation with people (parents) and organizations within the school context. 7. Competence in reflection and development, referring to the ability to reflect on one's own competence and to keep up with changing demands and developments within the profession. This set of requirements will be the basis for many assessment instruments at Dutch teacher colleges.

### *3.2.1.2 Performance criteria*

Having described the domains of competence, an important question is how to formulate criteria against which to judge teacher performances. Following our model, a comprehensive approach in formulating criteria is desirable to prevent an overreliance on isolated teacher activities, separate knowledge aspects, or on students' results. Instead, these elements of competent performance should be combined within verbal descriptors of criteria. Examples of one-sided, sometimes tautological criteria can easily be found:

'the teacher indicates clearly', 'chooses material in the correct way'. Following our model of competence (see Figure 1) performance criteria start with desirable learning activities and outcomes among students, from which acceptable teacher actions and decisions can be derived. The acceptability of teacher decisions has to do with the quality of the professional knowledge base with respect to specific teaching situations. Frederiksen, Sipusic, and Sherin and Wolfe (1998) speak of 'functional criteria'. An example of a functional criterion is taken from the context of instructional competence of kindergarten teachers, aimed at concept acquisition of young children, is presented by Roelofs and Van den Berg (2005): "by means of instructional activities (questions, explanations, performance tasks, discussions) the teacher succeeds accomplishing that children perform activities which contribute to a deeper understanding of a chosen set of concepts (e.g. autumn)."

### *3.2.1.3 Levels of performance*

Whereas criteria can be considered as statements of competent performance within task situations, standards look at the quality of the actions and their results. For our purposes, a discussion of developing performance standards is beyond the scope of this paper. Instead we emphasize the importance of having an interpretive model which can describe and explain differences in levels of performance. A model that accounts for differences between novices and experts within a profession can add to the construct validity of teacher assessment. The work of Berliner (2001) in the domain of expertise development is particularly relevant here. Berliner summarizes how expert professionals differ from novices. Experts:

- (a) excel in their own specialist domain and in specific contexts,
- (b) develop automation of actions which occur often,
- (c) are more opportunistic and flexible,
- (d) are more sensitive to task requirements and situations when they solve problems,
- (e) represent problems in qualitatively different (richer) ways,
- (f) recognize patterns in work situations more quickly and more accurately,
- (g) observe more significant patterns in the domain in which they are experienced,
- (h) take a longer period of preparation for solving problems and use richer and more personal sources of information in doing so.

The Berliner features of expert teachers were used recently model to investigate the construct validity of the NBTPS certification system for advanced teachers (Bond, Smith, Baker, Hattie, 2000). The results indicated that on the vast majority of the dimensions certified teachers outperformed non-certified teachers.

### 3.2.2 Sources of evidence of competence

The focus in the development of instruments is on obtaining the best possible evidence of competence of a candidate. Taking into account validity requirements related to content representation, underlying theory and process models, and generalizability, 'best possible' refers to the representativeness of tasks and task situations and the degree to which the assumed processes and effects of competent performance are adequately represented in the assessment.

The first choice in collecting evidence of competence relates to the nature of the evidence. Basic forms of evidence are: lesson documentation, lesson observation (live or recorded, focus on teacher actions or student activities), teacher logs (focus on actions), reflective interview (focus on decision-making processes), reflective report (focus on decision-making processes), student tests (focus on results), written teacher test (focus on knowledge base or decision-making processes), multimedia teacher test (focus on knowledge base or decision-making processes).

Following our model of competence, all the evidence of competence should be registered and interpreted within specific teaching situations. Competence instruments differ sharply in this respect.

Lesson observations range, for example, from context-free assessments on the basis of visits to lessons ('teacher explains clearly') to narrative reports of lesson episodes or unfiltered video sequences. This also applies to the gathering of lesson documentation, which may, for example, concern materials from complete series of lessons or extracts from what teachers regard as their best work. The contents of documentation can also vary sharply. They may relate to the teacher's lesson plans, to (examples of) teacher feedback to students, or to (examples of) what the students pick up from feedback from the teacher. In all this, it must be determined who is in the best position to supply the necessary evidence of competence: the teacher himself or herself, colleagues, students, parents, managers, external experts, or others. Following our model of competence, every participant in the assessment should be in a position to give representative and convincing pieces of evidence of competence related to consequences, teacher actions and decision making processes. <sup>1</sup>

The second choice in collecting evidence of competence includes the selection of a set of tasks and task situations which can be considered as representative, both quantitatively and qualitatively for the domain of competence under study. The following questions need to be answered:

- (a) Is the chosen situation or set of situations representative of actions in the professional situation?
- (b) Is the chosen task or set of tasks and task situation or set of task situations relevant or critical for demonstrating competence?
- (c) How difficult/complex is the task or task situation?
- (d) Does the candidate have the opportunity to supply the necessary evidence of competence?
- (e) Does the chosen set of tasks and task situations cover the universe of tasks and task situations?
- (f) Can statements on the situations being measured be extrapolated to the work situation?

Tasks and task situations differ in their degree of authenticity: real, simplified real, simulated and symbolized assessment situations. Assessment in real situations means that the candidate carries out tasks which arise in day-to-day reality, without the assessors' intervention in this situation. An example of a real situation is giving a lesson to one's own group or supervising one's own students in carrying out independent tasks. All teaching tasks can arise and have to be carried out on the spot. Success or failure in performing the teaching tasks has direct consequences for the students.

In simplified real task situations the candidate carries out a real task, which however is less complex than in reality, such as a mini-lesson with a small group of students. In simulated work situations, the direct consequences for the students and the possibility of 'stopping' the work situation are also lacking. The authenticity of the evidence of competence is reduced in simplified real or simulated task situations compared with real working situations. The advantage, however, is that it is possible to present relevant tasks which do not often occur under work conditions. Competence can also be

assessed in symbolic lesson situations in which the situation does not actually arise and the time pressure and immediacy of the lesson situation are lacking. The emphasis is on collecting evidence about decisions in various kinds of task situations. Although the level of authenticity is low, good coverage of tasks and situations can be achieved in described lesson situations.

In general, developers of assessment instruments skip deliberations about the nature and extent of the evidence required and immediately start designing instruments. Consequently, the significance of the various sources of evidence may become unclear. This problem arises in the assessment of (unstructured) portfolios. The nature of the evidence gathering can vary dramatically in the case of portfolios. Portfolios may contain direct evidence in the form of lesson artifacts, student achievements and reflective reports, but also products which in themselves are the outcome of assessments, such as the results of written tests, letters of recommendation, and assessments by peers. A portfolio can be a very useful way of putting together various pieces of evidence of competence. However, the assessability is heavily dependent on the structure of the portfolio and the admissibility, observability, and scorability of the recorded evidence (Heller, Sheingold, & Myford, 1998).

#### **4. Discussion**

In this paper a comprehensive framework for teacher competence was presented that can form the basis for valid assessments. This section is started by drawing up some conclusions. After that possible advantages of the model are discussed when using it to set up interpretive arguments. The paper is concluded by discussing recent applications of the model in the context of video portfolio assessment.

The first part of the paper was devoted to the development of the model. Based on reviews of literature it was concluded that there is no sound and broad-based scientific framework of what constitutes competent teaching. Several different elements of teacher competence have been emphasized throughout the history of evaluating teachers: personality traits which help to make a successful teacher; essential knowledge elements involving subject matter content, teacher thinking within a discipline; forms of teacher behavior which contribute to learning performance; practical knowledge and subjective theories of teachers determining teachers' actions in specific teaching situations, and teaching as the facilitation of powerful learning activities among learners.

All separate elements of teaching cover some aspects of teacher competence, but none of them fully describe or explain what competent teaching is. Therefore a unified comprehensive concept of teaching competence was introduced that takes into account all of the different elements of teacher competence.

In summary, the model states that teaching competence is reflected in the consequences of teachers' actions, the most important being students' learning activities. Starting from the consequences, the remaining components of the model were mapped backwardly. The component 'actions' refers to professional activities that foster student learning or other consequences. The component 'decision making' means a teacher has to make many decisions, either on a long term basis or immediately within a classroom situation, e.g. whether or not to initiate certain actions. In addition it was emphasized that decision making, actions and consequences take place within a specific context in which the teacher carries out his professional tasks. Finally, when making decisions and performing activities, teachers will have to draw from a professional knowledge base and from some personal characteristics.

In the second part of the paper it was described how the model can aid the development of assessment domains, performance criteria and the collection of evidence of competence. In doing this, it was discussed how three of Messick's criteria for construct validity could be met: content, theory and process models, generalizability.

Following the model of competent performance criteria should start with desirable learning activities and outcomes among students, from which acceptable teacher actions and decisions can be derived. It was stated that increasingly domains of competence are being stated in this way, not favoring a certain line of action but describing broad categories of activities and elaborating desirable learning activities. In addition it was stated that to distinguish levels of desirable performance, systematic comparisons between novices and experts within a profession can be made.

It was concluded that the focus in the development of instruments is on obtaining the best possible evidence of competence of a candidate. Taking into account validity requirements related to content representation, underlying theory and process models, and generalizability, 'best possible' refers to

the representativeness of tasks and task situations and the degree to which the assumed processes and effects of competent performance are adequately represented in the assessment.

With respect to the selection of sources of evidence it was emphasized that this should depend on the degree to which the complete process of competent performance is represented, i.e. teacher decisions, teacher actions, student actions. Finally it was concluded that a set of tasks and task situations should be chosen which can be considered as representative, both quantitatively and qualitatively for the domain of competence under study.

In terms of Kane (1992), the model can be used to set up interpretive arguments to substantiate judgments about teacher competence. We agree with Kane that competence can hardly ever be proven. More likely, an interpretive argument about teacher competence can at best be plausible. If assessors are able to interpret assessment results in terms of the postulated processes of our model, the interpretive argument is supported. For example, assessors judging the quality of instruction may interpret student results in terms of the way teachers make decisions when giving instruction, how they act, and what the consequences are for students within a specific classroom environment.

An advantage of the model is that varying or changing views on teaching do not affect the structure of the model. Different views on teaching, say programmed instruction versus discovery learning, will somehow be reflected in some kind of desirable learning activities, a repertoire of adequate actions, and accompanying decision processes on the part of the teacher. Whatever one's view on teaching is, developers of assessments can base their data collection decisions more consciously on the processes they would like to elicit in their assessments.

The model presented may form the basis for further professional development on the part of the teachers as it describes the processes teachers engage in. These very processes can be changed and adapted when teachers receive feedback and engage into reflective activities. It may thus help improve the quality of the training and learning process of the teachers.

Starting from a comprehensive model of teacher competence touches upon the question to what extent all aspects of competent performance should be covered in one assessment task. In other words: if different tasks and different sources of evidence are used, how will they be combined into one judgment? In the light of this question, we would like to conclude with some findings of a recent study at the University of Leiden and at the Dutch Institute for Educational Measurement (CITO) in which our model is used as the basis for collecting evidence of competence. Using video portfolios, different pieces of evidence for competent performance related to the same set of teaching situations are collected in coherence. What is recorded is the teachers' actions (using video), the underlying decision-making processes (using interviews), the consequences observed for students (using video), and the lesson context (situation) in which the teaching actions were demonstrated (using lesson documents). A scoring system has been constructed by which experienced assessors arrive at overall judgments on instructional competence. Results of a pilot study (Roelofs & Van den Berg, 2004, 2005) show that the portfolios are considered as cohesive and assessable collections of evidence. However, assessors do not use all evidence for arriving at judgments. Assessors observed evidence on teacher decision-making and on the task context but did not use it for their judgments. But at the same time, when reporting their judgments to the teacher, they would discuss these sources of evidence to provide more comprehensive feedback to the teachers. In that sense, the general model of competence was used for giving an interpretive argument.

## References

- Beijaard, D., & Verloop, N. (1996). Assessing teachers' practical knowledge. *Studies in Educational Evaluation*, 22(3), 275-286.
- Bellon, J.J., Bellon, E.C., & Blank, M.A. (1992). Teaching from a research knowledge base. New York: Merrill/MacMillan.
- Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research*, 35, 463-482.
- Bond, L., Smith, T., Baker, W.K., & Hattie, J.A. (2000). *The certification system of the National Board for Professional Teaching Standards: A construct and consequential validity study*. Greensboro, NC: Center for Educational Research and evaluation. The University of North Carolina.
- Brophy, J., & Good, T.L. (1986). Teacher behavior and student achievement. In M.C. Wittrock (Ed.), *Handbook of Research on Teaching* (pp. 328-375, 3<sup>rd</sup> ed.). New York: MacMillan.
- Bruner, J.S. (1963). *The process of education*. Cambridge: Harvard University Press.
- Creemers (1992), *Effectieve instructie: een empirische bijdrage aan de verbetering van het onderwijs in de klas*, 's-Gravenhage: SVO.
- Danielson, Ch., & McGreal, T.L. (2000). *Teacher evaluation to enhance professional practice*. Alexandria, VA, Princeton, NJ: ASCD/Educational Testing Service (ETS).
- Darling-Hammond, L. (2000). Teacher Quality and Student Achievement: A review of state policy evidence. *Education policy analysis archives*, 8 (1), 1-51.
- Darling-Hammond, L. (2003). *The effects of initial teacher education on teacher quality*. Paper presented at the Australian Council for Educational Research (ACER) Research Conference 2003 (Building Teacher Quality: What does the research tell us?), Melbourne.
- Darling-Hammond, L., LaFors, J., & Snyder, J. (2001). Educating teachers for California's future. *Teacher Education Quarterly*(Winter).
- Darling-Hammond, L., Wise, A.E. and Pease S.R. (1983). Teacher evaluation in the organizational context: a review of the literature. *Review of educational research*, 53, 285-237.
- Department of Education and Training (2004). *Competency Framework for teachers*. East Perth Western Australia: Department of Education and Training.
- Duffy, T. M., Lowyck, J., & Jonassen, D. H. (Eds) (1993). *Designing environments for constructive learning*. New York: Springer Verlag.
- Seattle, WA: Center for the Study of Teaching and Policy, University of Washington.
- Doyle, W. (1983). Academic work. *Review of Educational Research*, 53, 159-199.
- Duke, D. L., & Stiggins, R. J. (1990). Beyond minimum competence: evaluation for professional development. In J. Millmann and L. Darling-Hammond (eds) *The New Handbook of Teacher Evaluation. Assessing elementary and secondary teachers*. (pp. 116-132). Newbury Park, London, New Delhi: SAGE Publications.
- Dwyer, C. A. (1994). Criteria for performance-based teacher assessments: validity, standards, and issues. *Journal of Personnel Evaluation in Education*, 8(2), 135-150.
- Dwyer, C. A. (1998). Psychometrics of Praxis III: Classroom Performance Assessments. *Journal of Personnel Evaluation in Education*, 12(2), 163-187.
- Frederiksen, J. R, Sipusic, M, Sherin, M., & Wolfe, E. W. (1998). Video portfolio assessment of teaching. *Educational Assessment*, 5(4), 225-298.

- Getzels J.W. & Jackson, P.W. (1963). The teacher's personality and characteristics. In N.L. Gage (ed.). *Handbook of research on education*. Chicago: Rand McNally.
- Good, T.L. and J. E. Brophy. (1986). *Educational Psychology*, 3rd Edition. NY: Longman.
- Haertel, E.H. (1988). Assessing the Teaching Function. *Applied Measurement in Education*, 1(1), 99-107
- Haertel, E.H. (1991). New forms of teacher assessment. In G.Grant (ed.), *Review of Research in Education*. (pp. 3-29). Washington: American Educational Research Association.
- Heller, J. I.; Sheingold, K., & Myford, C. M. (1998). Reasoning about evidence in portfolios: cognitive foundations for valid and reliable assessment. *Educational Assessment*, 5(1), 5-40.
- Kagan, D.M. (1990). Ways of evaluating teacher cognition: inferences concerning the Goldilocks principle. *Review of Educational Research*, 60(3), 419-469.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527-535.
- Latham, A.S., Gitomer, D.H., & Ziomek, R. (1999); What the Tests Tell Us about New Teachers. *Educational Leadership*, 56(8), 23-26
- McConnell, E. A. (2001). Competence vs. competency. *Nursing Management*. 32(5):14.
- Medley, D. M., Coker, H., & Soar, R. S. (1984). *Measurement-based evaluation of teacher performance*. New York: Longman.
- Meijer, P.C., Verloop, N., & Beijaard, D. (1999). Exploring language teachers' practical knowledge about teaching reading comprehension. *Teaching and Teacher Education*, 15(1), 59-84.
- Merriënboer, J.J.G. van, Klink, M.R. van der, & Hendriks, M. (2002). *Competenties: van complicaties tot compromis*. The Hague: Onderwijsraad (Education Council of the Netherlands).
- Messick, S. (1996). Validity in performance assessments. In G.W.Phillips (ed.), *Technical issues in large-scale performance assessments*. Washington DC: National Center for Education Statistics.
- Mulder, M. (2001). *Competentieontwikkeling in organisaties. Perspectieven en praktijk*. The Hague: Elsevier bedrijfsinformatie.
- Parry, S.B. (1996). The quest for competencies. *Training*, July, 48-56.
- Pelkmans, A. (1998). *Evaluatie van leraren en schoolleiders*. Wageningen: Leerstoelgroep Agrarische Onderwijskunde.
- Peterson, K.D. (2002). *Teacher evaluation: a comprehensive guide to new directions and practices*. Thousand Oaks, CA: Corwin Press.
- Peterson, D., Micceri, T., & Smith, B. O. (1985). Measurement of teacher performance: A study in instrument development. *Teaching and Teacher Education*, 1, 63-77.
- Reynolds, M. (1992). What is competent beginning teaching? A review of the literature. *Review of Educational Research*, 62(1), 1-35.
- Roelofs, E.C., & Van den Berg, E.. (2004). *Videodossiers: leerkrachtbekwaamheid in beeld*. Paper presented during the 2004 educational research seminar (onderwijsresearchdagen) at Utrecht.
- Roelofs, E.C., & Sanders, P.F. (2007). Towards a framework for assessing teacher competence. *European journal of vocational training*. 40, 123-139

- Roelofs, E.C., & Van den Berg, E.. (2005). *Images of teacher competence. Design characteristics of a multimedia design portfolio*. Paper presented at the annual meeting of the American Educational Research Association, Montreal.
- Rosenshine, B. and Furst, N.F. (1973). The use of direct observation to study teaching. In R.M.W. Travers (ed.), *Handbook of Research on Teaching* (2nd Edition). Chicago: Rand McNally.
- Sanders, W.L., Saxton, A.M., & Horn, S.P. (1997). The Tennessee Value-Added Assessment System (TVAAS): A quantitative, outcomes-based approach to educational assessment. In J. Millman, (ed.). *Grading Teachers, Grading Schools*, Thousand Oaks, CA: Corwin Press.
- Santiago, P. (2002). *Teacher Demand and Supply: Improving Teaching Quality and Addressing Teacher Shortages*. Paris: Organisation for Economic Co-operation and Development (OECD).
- Shepard, L.A. (1993). Evaluating test validity. In L.Darling-Hammond (Ed.), *Review of research in education*. (pp. 405-450). Washington: AERA.
- Shuell, T. J. (1993). Toward an integrated theory of teaching and learning. *Educational Psychologist*, 28, 291–311.
- Simon, A., & Boyer, E.G. (1974). *Mirrors for behavior. An anthology of classroom observation instruments*. Wyncote: communication Material Center.
- Simons, P.R. (1999). Competentiegerichte leeromgevingen in organisaties en hoger beroepsonderwijs. In K. Schlusmans, R. Slotman, S. Nagtegaal, G. and Kinkhorst (red). *Competentiegerichte leeromgevingen*. Utrecht: Lemma.
- Spencer, L.M., & Spencer, S.M. (1993). *Competence at work. Models for superior performance*. New York: Wiley & Sons.
- Stichting Beroepskwaliteit leraren en ander personeel. (2003). *Bekwaamheidseisen leraren*. The Hague: SBL.
- Stodolsky, S.S. (1990). Classroom observation. In J., Darling-Hammond, & L. Millmann (Eds.), *The new Handbook of Teacher Evaluation. Assessing elementary and secondary school teachers*. (pp. 175-190). Newbury Park, California: Corwin Press.
- Thijssen, J.G.L. (1998). Hindernissen voor competentie management. *Opleiding & Ontwikkeling*, 11(10), 17-23.
- Thijssen, J.G.L. (2001). Personele flexibiliteit in strategisch perspectief: de betekenis van employability voor organisaties en individu. In R.F. Poel & J.W.M. Kessels (Red.), *Human Resource Development. Organiseren van het leren*, 103-116. Alphen aan den Rijn: Samsom.
- Tom, A.R., & Valli, L. (1990). Professional knowledge for teachers. In R.W. Houston (Ed.), *Handbook of research on teacher education* (pp. 372-392). New York: MacMillan.
- Uhlenbeck, A. (2002). *The development of an assessment procedure for beginning teachers of English as a foreign language*. Leiden: ICLON Graduate School of Education.
- Verloop, N. (1988). Investigating teacher cognitions. *Journal of Curriculum Studies*, 20, 81-86.
- Verloop, N., Beijaard, D., & Van Driel, J H. (1998). Beoordeling van docenten. *Pedagogische Studiën*, 75(6), 171-187.
- Vermunt, J. & Verschaffel, L. (2000). Process oriented teaching. In P.R.J. Simons, J. van der Linden and T. Duffy. *New Learning* (pp. 209-225). Dordrecht: Kluwer Academic Publishers.
- Wright, S.P., Horn, S.P., & Sanders, W.L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 1(1) 57-67.

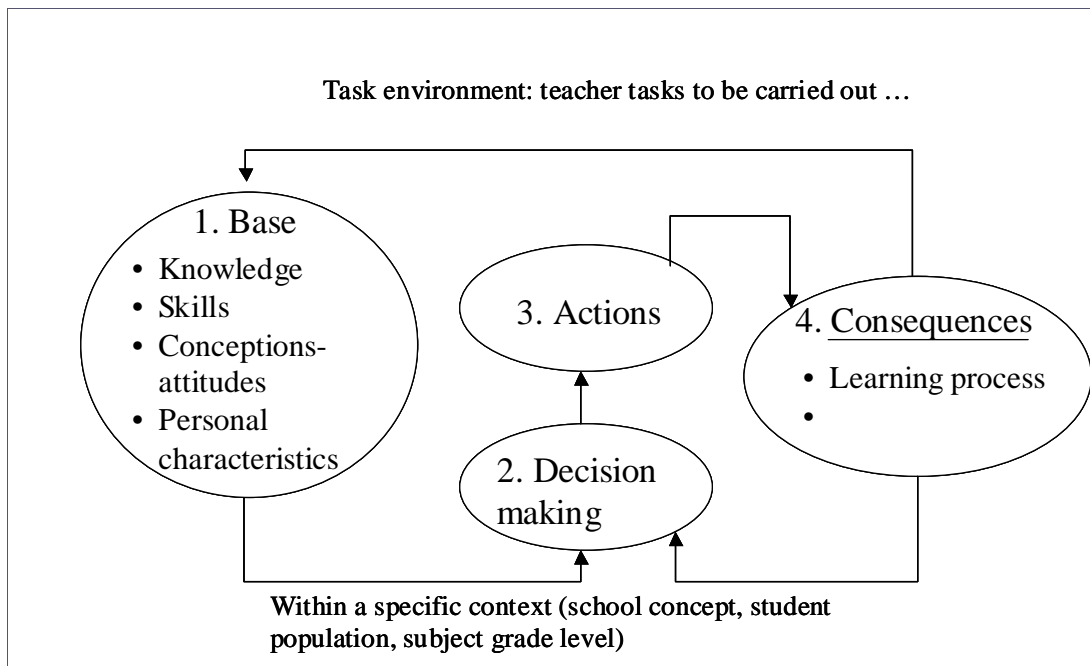


Figure 1: Interpretive model of competent performance (after Roelofs & Sanders, 2007)

---

<sup>i</sup>Peterson (2002) describes for different data sources the advantages and disadvantages of involving each of these participants.