Towards a framework for assessing teacher competence

Erik Roelofs

Senior researcher at the Psychometric Research and Information Centre of Cito, the Dutch National Institute of Test Development in Arnhem

Piet Sanders

Director of the Psychometric Research and Information Centre of Cito, the Dutch National Institute of Test Development in Arnhem

SUMMARY

Developing 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.

Based on contemporary insights into teaching and learning, an interpretive model of competent performance is described which, rather than being prescriptive in nature, offers scope for various forms of responsible professional performance. Consequences of professional performance for students/class/organisation 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 developing domains of competence and collecting evidence are discussed.

1. Introduction

Internationally, there is growing interest in assessing teacher competence prompted by demand for quality assurance and for greater recognition of the teaching profession (Verloop, 1999). The United States has a long tradition in teacher assessment, reflected in both the volume of research articles and books published, and the instruments developed. In the United States, the principle of accountability to taxpayers was a major incentive for directing attention at teacher assessment. Various instruments have been developed to assess teachers at various stages in their

professional careers in the context of selection, certification, and professional development (Dwyer, 1998). Much information in the present article was drawn from this body of literature. This does not, however, mean there is no teacher assessment tradition in countries beyond the United States. Pelkmans (1998), for instance, described teacher assessment practices in England, Wales, Germany, Australia, and the Netherlands. These countries do, however, have less experience than the United States.

In the Netherlands, as well as in other countries (Pelkmans 1998), teacher assessment is receiving greater attention because of the increased scope for policy-making by schools, one of the consequences of which is to make differentiation of position and pay possible (Verloop, 1999; Straetmans and Sanders, 2001). Growing emphasis on competence-based training is also increasing demand for assessing teacher competence. In addition, a law adopted in the Netherlands provides that professionals in education must satisfy competence requirements. The Dutch foundation for professional teaching competence (SBL, 2003) formulated requirements for seven domains of competence that are considered crucial for beginning teachers. In Section 3.2.1.1 these domains are further discussed.

The requirements must be met by teacher training colleges and school organisations. Finally, the option of allowing faster transfer into education from other occupations has been opened up to counteract the threat of teacher shortages in the Netherlands (Klarus, Schuler and Ter Wee, 2000; Tillema, 2001).

The developments described above call for a coherent approach to assessing teacher competence. This article presents some fundamentals for a framework for assessing teacher competence. We start by constructing an interpretive model for assessing competent (teacher) performance, based on different theoretical notions of good teaching. We go on to discuss various instruments for assessing teacher competence by using the requirements that follow from the concept of construct validity as expounded by Messick (1996). Lastly, we discuss some issues for future study.

2. An interpretive model of teacher competence

There is no generally accepted definition of the concept of competence. Recently several authors (e.g. Bos, 1998; Mulder, 2001; Van Merrienboer, Van der Klink and Jansen, 2002) have reviewed the literature and come 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 organisations, 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). Based

on a study of dozens of definitions of competence (e.g. Bunk, 1994; Spencer and Spencer, 1993; Parry, 1996), Mulder (2001) derived a definition that captures most of the important authors: 'competence is the ability of a person or organisation to achieve particular levels of performance' (p. 76). Citing different authors he adds that the competencies of individuals consist of:

- integrated action proficiencies
- which are made up of clusters of knowledge structures,
- cognitive, interactive, emotional, and where necessary psychomotor skills
- and attitudes and values which are necessary for:
- performing tasks,
- solving problems,
- and more generally the ability to function in a particular:
- occupation,
- organisation,
- position,
- role.

When measuring dimensions of competence, it must be noted that they are not directly observable, but are manifested in performance in a specific situation (Spencer and Spencer, 1993). In addition, competence can be developed to a particular level, for example beginner, advanced, and expert. Mulder (2001) emphasises 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 when gauging 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 for what constitutes competent teaching from which inferences can be drawn to assess 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 (see 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, 'programme-oriented' or 'development-oriented learning' in the context of early and pre-school education. The first approach is a logical one to adopt in the formulation of frameworks of competent performance which must apply to large groups of teachers, for example, national proficiency requirements. The second approach

is more appropriate for organisations 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 emphasised throughout the history of evaluating teachers. In reviewing the literature, different opinions of good teachers and good teaching can be distinguished (Creemers, 1991; Verloop, 1999):

(a) differentiating personality traits which help to make a successful teacher (Getzels and Jackson, 1963; Creemers (1991);

(b) describing knowledge elements involving subject matter content, ways teachers think within a discipline (Bruner, 1963; Tom and Valli, 1990);

(c) describing forms of teacher behaviour which contribute to learning performance (Brophy and Good, 1986; Simon and Boyer, 1974);

(d) describing teachers' cognition and decision-making processes (Kagan, 1990; Verloop, 1988);

(e) describing teachers' practical knowledge which they apply to specific situations in which they find themselves (their class, their subject domain) and the way they form theories about these situations (Beijaard and Verloop, 1996).

For each of these conceptions of good teaching, specific 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. A still dominant conception 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 (such as maths 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 and Blank, 1990). A frequently used method of assessing knowledge is taking standardised knowledge tests (e.g. Latham, Gitomer and Ziomek, 1999).

As a reaction to both emphasis on stable teacher characteristics and a one-sided emphasis on knowledge, teaching is also seen as displaying effective behaviour. This conception considers what a teacher shows in the classroom. Numerous observational instruments have been developed to concentrate on (small) units of behaviour, thought to be connected with successful learning outcomes (Stodolsky, 1990). However, this approach pays little attention to what goes on in the mind of teachers. What do they think or decide, and why do they decide the way they do? Different assessment instruments have been used to uncover teacher thinking, including thinking aloud protocols while solving a teaching problem and stim-

ulated recall interviews. During such interviews, teachers look back at their videotaped performance and answer questions on 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 confronted by teachers. 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 and Beijaard, 1999).

With the recent revolution in thinking about learning, summed up in the term 'new learning', there is an additional conception of teaching: promoting powerful learning activities among learners. In this conception, good teaching assumes that teachers do not so much need to demonstrate a clear-cut repertoire of 'correct' behaviour but rather show that they contribute to successful learning of their students (Simons, 1999; Vermunt and Verschaffel, 2000).

Although all the separate conceptions of teaching cover some aspects of teacher competence, none of them fully describe or explain what competent teaching is. Looking back at the definition of competencies described above, there is a need for a unified comprehensive concept of teaching competence that considers all the different elements of teacher competence, that is, teacher traits, teacher knowledge, teacher behaviour, teacher thinking, situation-specific decision-making, and resulting learning activities. Roelofs and Sanders (2003) have developed a model of competent performance for assessing teacher competence which does justice to the aspects of competent performance described above. The model follows the general definition of competencies described by Mulder (2001).

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 backwards. 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 (see 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 organisational conditions). Third, when making decisions and performing activities, teachers will have to draw from a professional knowledge base and from some personal characteristics.

Evaluating different domains of teacher competence, for example, instruction and classroom management, means that interpretive infer-

Figure 1: Interpretive model of competent performance
(based on Roelofs and Sanders, 2003)
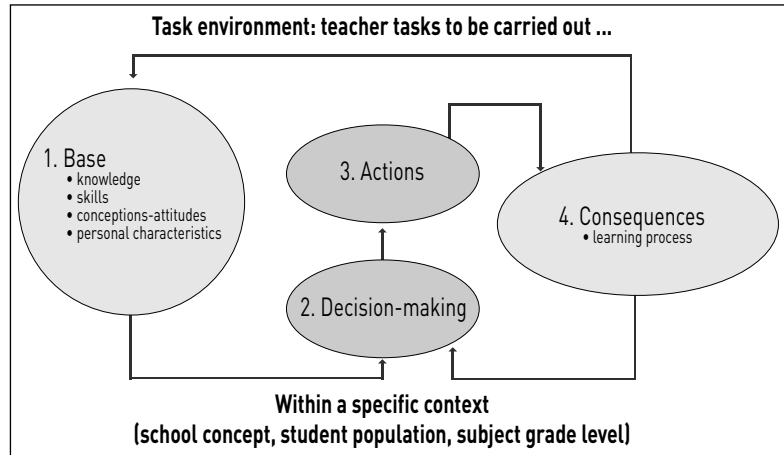
ences are made about teachers (see Kane, 1992). When combining different aspects of teaching into one comprehensive model of competent performance, the chances for valid inferences are better than when using reductionistic models which concentrate on separate parts of the teaching process.

3. Assessment of teacher competence

Before describing our general model's implications 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 determining the quality of instruments has been developed by Messick (1996). Messick indicates that for each form of assessment it is necessary to consider six aspects of construct validity:

(a) content,

(b) theory and process models,

(c) structure,

(d) generalisability,

(e) external aspects,

(f) consequences.

The 'content' aspect is about the relevance and representativeness of the assessment. The question is: within what limits can conclusions be drawn from the assessment? The 'theory and process models' aspect is about 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 minimised. The 'structural aspect' relates to whether the performance criteria correctly reflect the criteria that experts use and the accuracy and consistency with which performance is scored and assessed.

'Generalisability' is the extent to which assessments can be generalised to a universe of, for example, tasks and settings.

The external aspects of validity are the extent to which the measurement results converge with, and diverge from, other measurements and constructs. '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 considering 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 generalisability.

3.2.1. Developing 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 practising professionals (see Verloop, Beijaard and Van Driel, 1998). The mix of scientific and practical perspectives contributes to the acceptance and practical usability of instruments (Beijaard and Verloop, 1996; Duke and Stiggins, 1990; Uhlenbeck, 2002).

The basis of any assessment must contain 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

To demarcate domains of competence, selecting what is characteristic of adequate professional functioning and what is critical to functioning is necessary. 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 both in the US and the Netherlands. Danielson and McGreal (2000) distinguished four broad, relevant professional task areas: planning and preparation; instruction; classroom environment; professional responsibilities. Referring to the 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' (op. cit., p. 31). On 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' (op. cit., p. 32). These statements illustrate focus on the consequences of the actions, rather than on the actions themselves.

The description given by Danielson and McGreal 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, teachers' in-class practice in each of the four areas mentioned. In the Praxis-III assessment, classroom observations of teacher and student behaviour 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 certifying (advanced) teachers to dimensions of teaching expertise, such as improvisation, degree of challenge, passion for teaching and learning (Bond et al., 2000).

In the Netherlands, the Dutch Foundation for Professional Teaching Competence (SBL), recently developed a set of initial proficiency requirements for teachers in primary and secondary education for the Ministry of Education based on seven broad domains of competence. In the description of requirements, SBL starts with how classes and individual students of competent teachers function, illustrating a coherent approach to teaching. The following domains have been developed:

(a) interpersonal competence, the ability to create a friendly, cooperative climate and open communication;

(b) pedagogic competence, the ability to create a psychologically safe learn-
ing environment for students, contributing to their wellbeing;
(c) subject matter and didactic competence, the ability to guide students
in acquiring the basics of school subjects and the way this knowledge
can be used in daily and working life;
(d) organisational competence, the ability to create an orderly on-task cli-
mate in their classes;
(e) competence in cooperating with colleagues, the ability to gear one's
own work to the work of colleagues and to contribute to the school or-
ganisation in general;
(f) competence in cooperating with the school environment, the ability to
contribute to cooperation with people (parents) and organisations with-
in the school context;
(g) competence in reflection and development, the ability to reflect on one's
own competence and to keep up with changing demands and devel-
opments within the profession.

This set of requirements will be the basis for many assessment instru-
ments at 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 performance.
Following our model, a comprehensive approach in formulating criteria is
desirable to prevent 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 easi-
ly 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 ac-
ceptability of teacher decisions has to do with the quality of the profession-
al knowledge base with respect to specific teaching situations. Frederik-
sen et al. (1998) speak of 'functional criteria'. An example of a functional
criterion, taken from the instructional competence of kindergarten teach-
ers aimed at concept acquisition of young children, is presented by Roelofs
and Van den Berg (2005): 'by means of instructional activities (ques-
tions, explanations, performance tasks, discussions) the teacher succeeds
when children perform activities which contribute to a deeper understand-
ing of a chosen set of concepts (e.g. autumn)'.

3.2.1.3. Levels of performance

Where criteria can be considered statements of competent perform-
ance within task situations, standards look at the quality of the actions and
their results. For our purposes, a discussion on developing performance

standards is beyond the scope of this article. Instead we emphasise 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 significant. Berliner summarises 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) recognise 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.

The Berliner features of expert teachers were used recently model to investigate the construct validity of the NBTPS certification system for advanced teachers (Bond et al., 2000). Results showed that in most dimensions certified teachers outperformed non-certified teachers.

3.2.2. Sources of evidence of competence

In developing instruments, focus is on obtaining the best possible evidence of competence in a candidate. Considering validity requirements related to content representation, underlying theory and process models, and generalisability, '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 based on visits to lessons ('teacher explains clearly') to narrative reports of lesson episodes or unfiltered video sequences. This also applies to gathering lesson documentation, which may, for example, concern ma-

terials 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 teachers' lesson plans, to (examples of) teacher feedback to students, or to (examples of) what the students pick up from feedback from teachers. In all this, it must be determined who is in the best position to supply the necessary evidence of competence: teachers themselves, 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. Peterson (2002) describes for different data sources the advantages and disadvantages of involving each of these participants.

The second choice in collecting evidence of competence includes selecting 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 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 symbolised assessment situations. Assessment in real situations means that the candidate carries out tasks which arise in day-to-day reality, without 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 teaching tasks has direct consequences for 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 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 to real working situations. The advantage, however, is that it is possible to present relevant tasks which do not often occur in 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. Emphasis is on collecting evidence about decisions in various 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 assessing (un-structured) portfolios. The nature of 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 very useful for 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 and Myford, 1998).

4. Discussion

This article presents a comprehensive framework for teacher competence that can form the basis for valid assessments. This section starts by drawing up some conclusions. It then discusses possible advantages of the model when using it to set up interpretive arguments. It concludes by looking at recent applications of the model in video portfolio assessment.

The first part of the article is devoted to developing the model. Based on reviews of literature it concluded there is no sound and broad-based scientific framework of what constitutes competent teaching. Several different elements of teacher competence have been emphasised 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 behaviour which contribute to learning performance; practical knowledge and subjective theories of teachers determining teachers´ actions in specific teaching situations, and teaching as promoting 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 to consider all 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 in reverse. The component 'actions' re-

fer to professional activities that foster student learning or other consequences. The component 'decision-making' means a teacher has to make many decisions, either long-term or immediately in a classroom situation, e.g. whether or not to initiate certain actions. In addition, it was emphasised that decision-making, actions and consequences take place within a specific context in which teachers carry out their professional tasks. Finally, when making decisions and performing activities, teachers will have to draw from a professional knowledge base and from some personal characteristics.

The second part of the article describes how the model can help develop assessment domains, performance criteria and collect evidence of competence. It also discusses how three of Messick's criteria for construct validity could be met: content, theory and process models, generalisability.

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. Increasingly, domains of competence are being stated in this way, not favouring a certain line of action but describing broad categories of activities and developing desirable learning activities. In addition, to distinguish levels of desirable performance, systematic comparisons between novices and experts within a profession can be made.

The focus for developing instruments is on obtaining the best possible evidence of competence of candidates. Considering validity requirements related to content representation, underlying theory and process models, and generalisability, '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.

On selecting sources of evidence, 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, a set of tasks and task situations should be chosen which can be considered representative, both quantitatively and qualitatively for the domain of competence under study.

According to Kane (1992), the model can be used to set up interpretive arguments to substantiate judgements on 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 the way teachers make decisions when giving instruction, how they act, and what the consequences are for students in a specific classroom environment.

An advantage of the model is that varying or changing views on teach-

ing do not affect itse structure. Different views on teaching, such as 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 the view on teaching, assessment developers 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 teachers as it describes the processes teachers engage in. These processes can be changed and adapted when teachers receive feedback and engage in reflective activities. It may thus help improve the quality of training and learning process of teachers.

Starting with a comprehensive model of teacher competence raises 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 judgement? We would like to conclude with some findings of a recent study at the University of Leiden and the Dutch Institute for Educational Measurement (CITO) in which our model was 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 were collected coherently. Teachers' actions (on video), underlying decision-making processes (by interview), consequences observed for students (on video), and the lesson context (situation) in which the teaching actions were demonstrated (using lesson documents) were recorded. A scoring system was constructed through which experienced assessors arrive at overall judgements on instructional competence. The first results of a pilot study (Roelofs and Van den Berg, 2004, 2005) show that portfolios as cohesive and assessable collections of evidence. However, assessors do not use all evidence for arriving at judgements. Assessors had evidence on teacher decision-making and on the task context but did not use it for their judgements. But when reporting their judgements to teachers, they discussed these sources of evidence to provide more comprehensive feedback to teachers. This is how the general model of competence was used for giving an interpretive argument.

Bibliography

Beijaard, D.; Verloop, N. Assessing teachers' practical knowledge. *Studies in Educational Evaluation*, Vol. 22, No 3, 1996, p. 275-286.

Bellon, J. J.; Bellon, E. C.; Blank; M.A. *Teaching from a research knowledge base*. New York: MacMillan, 1992.

Berliner, D. C. Learning about and learning from expert teachers. *International Journal of Educational Research*, Vol. 35, 2001, p. 463-482.

Bond, L. et al. *The certification system of the National Board for Professional Teaching Standards: a construct and consequential validity study.* Greensboro, NC: University of North Carolina, Center for Educational Research and Evaluation, 2000.

Brophy, J.; Good, T. L. Teacher behaviour and student achievement. In Wittrock, M.C. (ed.). *Handbook of research on teaching*, 3rd ed., p. 328375. New York: MacMillan, 1986.

Bruner, J. S. *The process of education.* Cambridge: Harvard University Press, 1963.

Creemers, H. P. M. *Effectieve instructie: een empirische bijdrage aan de verbetering van het onderwijs in de klas.* 's-Gravenhage: Instituut voor Onderzoek van het Onderwijs, 1992.

Danielson, Ch.; McGreal, T.L. *Teacher evaluation to enhance professional practice.* Alexandria, VA: Association for Supervision & Curriculum Development, 2000.

Doyle, W. Academic work. *Review of Educational Research*, Vol. 53, 1983, p. 159-199.

Duke, D. L.; Stiggins, R.J. Beyond minimum competence: evaluation for professional development. In Millmann J.; Darling-Hammond L. (eds.). *The new handbook of teacher evaluation. Assessing elementary and secondary teachers,* p. 116-132. Newbury Park, CA: Sage, 1990.

Dwyer, C. A. Criteria for performance-based teacher assessments: validity, standards, and issues. *Journal of Personnel Evaluation in Education,* Vol. 8, No 2, 1994, p. 135-150.

Dwyer, C. A. Psychometrics of Praxis III: Classroom performance assessments. *Journal of Personnel Evaluation in Education,* Vol. 12, No 2, 1998, p. 163-187.

Frederiksen, J. R. et al. Video portfolio assessment of teaching. *Educational Assessment,* Vol. 5, No 4, 1998, p. 225-298.

Getzels J. W.; Jackson, P.W. The teacher's personality and characteristics. In Gage, N.L. (ed.). *Handbook of research on education.* Chicago: Rand McNally, 1963.

Haertel, E. H. New forms of teacher assessment. In Grant, G. (ed.). *Review of research in education.* Washington, DC: American Educational Research Association, 1991, p. 3-29.

Heller, J. I.; Sheingold, K.; Myford, C.M. Reasoning about evidence in portfolios: cognitive foundations for valid and reliable assessment. *Educational Assessment,* Vol. 5, No 1, 1998, p. 5-40.

Kagan, D. M. Ways of evaluating teacher cognition: inferences concerning the Goldilocks principle. *Review of Educational Research*, Vol. 60, No 3, 1990, p. 419-469.

Kane, M.T. An argument-based approach to validity. *Psychological Bulletin*, Vol. 112, No 3, 1992, p. 527-535.

Klarus, R.; Schuler,Y.; Ter Wee, E. *Bewezen geschikt: een assessmentprocedure voor aspirant leerkrachten (zij-instromers) in het PO en VO.* Wageningen: Stoas Onderzoek, 2000.

Latham, A.S.; Gitomer, D.H.; Ziomek, R. What the tests tell us about new teachers. *Educational Leadership*, Vol. 56, No 8, 1999, p. 23-26.

McConnell, E. A. Competence vs. competency. *Nursing Management*, Vol. 32, No 5, 2001, p. 14.

Meijer, P. C.; Verloop, N.; Beijaard, D. Exploring language teachers' practical knowledge about teaching reading comprehension. *Teaching and Teacher Education,* Vol. 15, No 1, 1999, p. 59-84.

Merriënboer, J. J. G. van,; Klink, M. R. van der; Hendriks, M. *Competenties: van complicaties tot compromis.* The Hague: Onderwijsraad (Education Council of the Netherlands), 2002.

Messick, S. Validity in performance assessments. In Phillips, G.W. (ed.). *Technical issues in large-scale performance assessments.* Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, 1996.

Mulder, M. *Competentieontwikkeling in organisaties: perspectieven en praktijk.* The Hague: Elsevier bedrijfsinformatie, 2001.

Parry, S.B. The quest for competencies. *Training*, July 1996, p. 48-56.

Pelkmans, A. *Evaluatie van leraren en schoolleiders.* Wageningen: Leerstoelgroep Agrarische Onderwijskunde, Landbouwuniversiteit Wageningen, 1998.

Peterson, K.D. *Teacher evaluation: a comprehensive guide to new directions and practices.* Thousand Oaks, CA: Corwin Press, 2002.

Reynolds, M. What is competent beginning teaching? A review of the literature. *Review of Educational Research,* Vol. 62, No 1, 1992, p. 1-35.

Roelofs, E. C.; Van den Berg, E. *Videodossiers: leerkrachtbekwaamheid in beeld.* Paper presented during the 2004 educational research seminar (onderwijsresearchdagen) at Utrecht, 2004.

Roelofs, E. C.; Sanders, P. Beoordeling van docentcompetenties. In Mulder, M. et al. (eds.). *Bevoegd, maar ook bekwaam?* Groningen: Wolters-Noordhoff, 2003, p. 277-299.

Roelofs, E. C.; Van den Berg, E. *Images of teacher competence. Design characteristics of a multimedia design portfolio.* Paper presented at the annual meeting of the American Educational Research Association, Montreal, 2005.

Shepard, L. A. Evaluating test validity. In Darling-Hammond, L. (ed.). *Review of research in education.* Washington, DC: American Educational Research Association, 1993, p. 405-450.

Simon, A.; Boyer, E. G. *Mirrors for behaviour: an anthology of classroom observation instruments.* Wyncote, PA: Communication Material Center, 1974.

Simons, P. R. Competentiegerichte leeromgevingen in organisaties en hoger beroepsonderwijs. In Schlusmans, K. et al. (eds.). *Competentiegerichte leeromgevingen.* Utrecht: Lemma, 1999.

Spencer, L.M.; Spencer, S.M. Competence at work: models for superior performance. New York: Wiley, 1993.

Stichting Beroepskwaliteit leraren en ander personeel. *Bekwaamheidseisen*

leraren. The Hague: Stichting Beroepskwaliteit Leraren en ander onder-
wijspersoneel, 2003.

Stodolsky, S.S. Classroom observation. In Darling-Hammond, J.; Millmann,
L. (eds.). *The new handbook of teacher evaluation: assessing elemen-
tary and secondary school teachers*, p. 175-190. Newbury Park, CA:
Sage, 1990.

Straetmans, G.J.J.M.; Sanders, P.F. *Beoordelen van competenties van
docenten.* Utrecht: Programmamanagement Educatief Partnerschap,
2001.

Thijssen, J.G.L. Hindernissen voor competentiemanagement. *Opleiding &
Ontwikkeling,* Vol. 11, No 10, 1998, p. 17-23.

Thijssen, J.G.L. Personele flexibiliteit in strategisch perspectief: de beteke-
nis van employability voor organisaties en individu. In Poel, R.F.; Kessels,
J.W.M. (eds.). *Human resource development: Organiseren van het leren,*
p. 103-116. Alphen aan den Rijn: Samsom, 2001.

Tillema, H. *Assessment van competenties: van beoordelen naar ontwikke-
len .* Alphen aan den Rijn: Kluwer, 2001.

Tom, A.R.; Valli, L. Professional knowledge for teachers. In Houston R.W.
(ed.). *Handbook of research on teacher education.* New York: MacMil-
lan, 1990, p. 372-392.

Uhlenbeck, A. *The development of an assessment procedure for begin-
ning teachers of English as a foreign language.* Leiden: ICLON Grad-
uate School of Education, 2002.

Verloop, N. Investigating teacher cognitions. *Journal of Curriculum Stud-
ies*, Vol. 20, 1988, p. 81-86.

Verloop, N. *De leraar: reviewstudie uitgevoerd in opdracht van PROO.* Den
Haag: NWO/PROO, 1999.

Verloop, N.; Beijaard, D.; Van Driel, J.H. Beoordeling van docenten.
Pedagogische Studiën, Vol. 75, No 6, 1998, p. 171-187.

Vermunt, J.; Verschaffel, L. Process oriented teaching. In Simons, P.R.J.;
van der Linden, J.; Duffy, T. *New learning.* Dordrecht: Kluwer, 2000,
p. 209-225.