Chapter 12
The Power of Assessment in Teacher Education

Judith Gulikers, Dominique Sluijsmans, Liesbeth Baartman and Paul Bartolo

Introduction

This chapter addresses three main assessment challenges faced in innovative assessment practices and aims to encourage teacher educators to take up these challenges in their assessment practices.

The *first challenge* is to establish a shift from de-contextualised tests to more authentic assessments. This requires the development of assessments that observe student teachers' performance in situations that resemble the current and future teaching practice as much as possible.

A *second challenge* is to increase student teachers' involvement in assessment by handing over the responsibilities of assessor from teacher educator to student teacher. After all, one of the main responsibilities of teachers is to assess their pupils in schools. For teacher educators, this means that they need to equip student teachers with assessor skills, like developing appropriate assessment criteria and giving adequate feedback on student performance.

Because assessment is a delicate issue and many important educational decisions are based on assessment outcomes, it is important to assure the quality of assessment. A *third challenge* is, therefore, that teacher educators critically evaluate the quality of their own assessments and equip student teachers with the skills to do so in their own future practice as well. This chapter addresses these three challenges and offers practical guidelines for dealing with them in daily assessment practices.

Power of Assessment in Teacher Education

It is widely recognised that the main goal of professional higher education is to help students to become 'reflective practitioners' who are able to reflect critically upon their own professional roles (Schön, 1987). Teacher educators have the responsibility to educate student teachers to be competent teachers, who reflect on their own practices and improve, develop and change constantly. To measure to what extent a

J. Gulikers
Wageningen University and Research Centre, Wageningen, the Netherlands

student teacher is a competent teacher and to help student teachers to reflect on their own practice and support their development, high-quality assessment is crucial. This chapter addresses new models of assessment that give rise to innovative assessment practices.

In the last few decades, educational testing practices changed from what is referred to as 'the testing culture' to the 'assessment culture'. The testing culture is characterised by standardised tests, mostly of a multiple-choice format, that mainly address factual knowledge or routine-based skills. Tests had a *summative* function, meaning that they were conducted at the end of an instructional period to test and judge if the students had learned what was presented in lectures or books, for the purpose of certifying or grading students. A main quality criterion for tests was that they needed to give a reliable score of the 'true' knowledge level of the student. Birenbaum (1996) characterised educational practices during the testing culture by (1) knowledge transmission as the main instructional method, (2) rote learning and (3) summative and standardised testing.

These educational practices, however, did not stimulate students to develop competences required for the changing labour market demands. Students were not equipped with the necessary competences to be flexible in the changing world and to continuously adapt and develop their own (teaching) practices. It was expected that assessment could play a crucial role in preparing student teachers to become flexible and reflective teachers, as a growing body of empirical evidence showed that assessment is one of the main driving forces behind student learning and competence development (e.g. Gibbs, 1992). This implies changing ideas about the function of assessment in teacher education. Next to summative assessment of learning, assessment should also be used during the learning process as assessment for learning. This means that assessment is used to diagnose the current level of competence and to give feedback on this current performance in order to stimulate further development towards becoming a professional teacher (Elwood & Klenowski, 2002). When assessment is used for this purpose, it is called *formative assessment.*

Educational practices in the assessment culture are characterised by (1) instruction that aims at stimulating student learning; (2) learning based on active knowledge construction and (3) both formative and summative assessment in the form of contextualised, performance-based assessments that address professional competence development (Birenbaum, 1996, 2003). This shows that 'assessment' means much more than only measuring and judging; it should play a crucial role in the whole learning process. More emphasis is placed on congruence between instruction and assessment, which should both focus on stimulating the development of competences needed to flexibly perform various professional roles and on stimulating reflection and lifelong learning skills by involving students as active participants in the learning process.

In response to the assessment culture, portfolio assessment was introduced in teacher education as a powerful tool to promote and support continuous monitoring of students' competence development (Klenowski, 2002). In fact, the use of electronic portfolios has become a new trend in teacher education. In a portfolio, student teachers collect evidence of their learning process and/or competence levels, during

several time intervals and through different kinds of assessment tasks. The evidence is often organised around specific competences and may be supplemented with reflections on educational achievement and on personal and professional development. Portfolios were primarily introduced to collect and assess performances in authentic contexts and to encourage learners to reflect on their performances (Järvinen & Kohonen, 1995). This chapter does not elaborate on portfolio assessment, however, the assessments described here can all be part of a portfolio assessment. The first challenge, for example, deals with developing assessments to assess authentic performance often incorporated in portfolios.

Challenge I: Towards Authentic Assessments

The need to change assessments in interesting, authentic and contextualised tasks is described as one of the main challenges in education (Birenbaum, 1996). Authentic assessments are based in professional practice as much as possible and confront students with situations that require them to demonstrate the competences professionals would use in the same situation in their daily practice (Gulikers, Bastiaens, & Kirschner, 2004).

In teacher education, the importance of authentic assessment has been recognised for quite some time (e.g. Darling-Hammond & Snyder, 2000). This resulted, for example, in an increase of assessments conducted at the workplace, the schools. However, these assessments require thoughtful planning. Teacher educators tend to develop authentic assessments based on their own ideas of what professional practice entails, instead of on thorough knowledge of authentic assessment. Cooper (1994) describes an example in which mathematics teachers thought they developed a very realistic mathematics assessment, while students perceived the assessments as artificial and fake and they experienced the assessment as confusing. As a result, the assessment hampered their learning. In other words, an authentic assessment must not only be more realistic in the eyes of the teacher educator, but the student teachers must experience the assessments to be relevant and representative of their future professional work (Gulikers, Bastiaens & Kirschner, 2004; McDowell, 1995). What makes this even more complicated is that the perception of student teachers about what is, or is not, an authentic assessment depends on the extent and nature of their practical work experience. This implies that some kinds of authentic assessments might be more useful in particular stages of the teacher education programme than others.

The following section describes a framework for the development or evaluation of assessments with different degrees of authenticity and offers guidelines for the way in which this framework can be applied in teacher education.

Five-Dimensional Framework for Assessment Authenticity, 5DF

Gulikers et al. (2004) and Gulikers, Bastiaens and Kirschner (2006) developed a five-dimensional framework (5DF) that describes five main assessment qualities that influence assessment authenticity, namely:

1. *Task.* The assessment assignment that defines the content of the assessment, e.g. using active teaching methods in your teaching that stimulates pupils to actively participate in class.
2. *Physical context.* The environment in which student teachers have to perform the assessment task, e.g. an assessment conducted in the teacher education institute, in a simulated setting, or during their internship in a classroom with 30 pupils.
3. *Social context.* The interaction (im)possibilities during the assessment, e.g. (im)possibilities to ask for assistance when a pupil gets really aggressive.
4. *Form.* The assessment method, independent of the content, e.g. assessing the task 'using active teaching methods in your teaching' with a written multiple-choice or open-answer test compared to a performance assessment in which students have to demonstrate their use of an active teaching method.
5. *Criteria.* The characteristics of the performance product/process that are valued, e.g. are the pupils paying attention? Did the pupils show signs of learning? Did the student teacher use more than one active teaching method successfully?

Changing from traditional tests to assessments that resemble teaching practice in all possible ways is a major challenge. The rationale behind this framework is that there is a range of possibilities between 'completely authentic' and 'completely in-authentic' assessment and that there are several ways to increase the authenticity of an assessment. The five dimensions of the framework reflect a realistic teaching situation to a more or lesser degree. From an inauthentic perspective, the task 'using active teaching methods in your teaching' can be assessed through an open-ended answer test (form) conducted individually (social context) in the teacher education institute (physical context) asking students to describe three concrete examples of their use of active teaching methods that are evaluated against criteria developed by the lecturer (criteria). From a more authentic perspective, this same task can be assessed through a performance assessment (form) conducted during their internship (the teaching practice of student teachers) in the classroom (physical context) in which student teachers have to actively involve the pupils (social context) by their teaching method and are assessed by criteria developed by their mentor at school (criteria).

Thinking about authentic assessment as a means of assessment that has several dimensions gives teacher educators tools to develop various kinds of authentic assessments, both for assessing student teachers' learning in teaching practice and at the teacher education institute, for example through realistic case-based assignments or project work based on professional problem situations. The 5DF legitimises authentic assessments for both internal and external quality assurance, without arguing that all assessments should strive for maximum authenticity.

Besides giving tools to develop and evaluate authentic assessments, this framework also supports the development of various authentic learning situations. As argued at the beginning of this chapter, the assessment culture places more emphasis on integrating teaching and assessment. A framework like 5DF supports the development of authentic assessments and helps to increase the congruence between assessment and instruction.

Practical Implications: How to Use Authentic Assessments?

Based on research that examined student perceptions of the five dimensions of the 5DF, several guidelines were developed for using the 5DF to construct authentic assessments during different phases of a learning trajectory.

General guidelines

- Confront students with authentic assessment early in their educational trajectory.
- Explicitly communicate the authenticity of a certain assessment and create mutual understanding between involved stakeholders, like teacher educators, student teachers and mentors.
- Use the 5DF in teams of teacher educators to explicate and discuss ideas concerning authentic assessments, as well as to develop and evaluate authentic assessments.

Rules of thumb concerning several authenticity dimensions

- Integrate instruction and assessment by offering opportunities to perform authentic, integrated tasks (i.e. learning tasks/formative assessments) in and out of school to prepare student teachers for summative authentic assessment.
- Stimulate teacher educators to keep up-to-date with developments and requirements in professional practice.
- Allow student teachers to tailor the assessment task and criteria to their own situations, like work context, interest and learning goals.

Rules of thumb concerning specific authenticity dimensions

- Do not make the assessment *task* completely authentic *for* student teachers, but help them to make the task authentic for themselves.
- Increase the authenticity of the *physical context* as student teachers gain more experience with working or assessing in practice.
- Developing an authentic social context is less important than the other four dimensions.
- When considering an authentic *social context*, first deal with traditional beliefs about individualistic or collaborative assessment.
- An authentic assessment *form* should involve multiple assessment methods and moments for different aspects of job performance.
- Consider incorporating knowledge testing directed at *knowing why* as part of the authentic assessment *form*.
- Involve teaching practice in the development and interpretation of authentic assessment *criteria*.
- Authentic assessment *criteria* should deal with *what* is done in teaching practice, as well as with *how* this is done. Criteria should change from being specific and step-by-step to being more open and allowing more student interpretation as students gain more experience with performing in practice.

Challenge II: Towards Student Teachers as Assessors

Besides a shift from more knowledge-based and de-contextualised tests towards more authentic assessments, a shift also occurred from teacher-directed assessment to a perspective in which students are given more responsibility in their assessment process. This shift fits with the need for student teachers to become self-directed professionals, who are able to continuously self-assess their performance (Boud, 1995). This self-directed assessment can be stimulated through the use of portfolio assessment in which student teachers often have the primary responsibility in the selection of the appropriate evidence and in reflecting on its relevance for their professional development (Klenowski, 2002).

Another tool for the development of self-directed assessment skills is peer assessment, which is the process whereby individuals evaluate the performances of their peer(s) and provide feedback on these performances (Freeman, 1995). Peer assessment can be introduced as a valuable tool to stimulate student teachers to critically evaluate their own performances as well as those of colleagues and their future pupils. Peer assessment is mainly used as a formative assessment tool, aimed at stimulating professional development and giving constructive feedback in order to stimulate further improvement.

There are various reasons why peer assessment is important for teacher education. First, the importance of communication between teachers in schools has been endorsed by many researchers. Teachers have to collaborate, learn from each other and become a member of a learning organisation (Verloop & Wubbels, 2000). Second, it is advisable to support student teachers to learn how to critically assess the performance of peers, as they will have to play the role of assessor in their future classrooms as well. A third reason is that after graduation, student teachers are likely to rely on the judgement of their colleagues in the school to evaluate the effectiveness of their teaching performance.

Thus, being able to interpret the work of colleagues and giving constructive feedback on these performances are necessary prerequisites for teachers' professional development and for improving their own functioning (Verloop & Wubbels, 2000). However, assessing the work of peers is a skill that needs to be developed (Birenbaum, 1996; Sluijsmans, Moerkerke, Dochy & Merriënboer, 2001). To understand the use of peer assessment and ways to teach this type of skills, the peer assessment skill was unravelled in several constituent skills (Sluijsmans, Brand-Gruwel, Merriënboer & Martens, 2004).

As shown in Fig. 12.1, the three main skills for peer assessment are (1) defining assessment criteria, (2) judging the performance of a peer and (3) providing feedback for future learning, for example giving constructive feedback about the product of a peer. A training programme for peer assessment should be based on these skills. The design of such a training programme results in a number of peer assessment exercises, which should be embedded in an existing course. The peer assessment exercises have a formative function in that they are aiming at helping student teachers develop the skill of assessing each other's performances to give feedback about how the development of these skills can be supported.
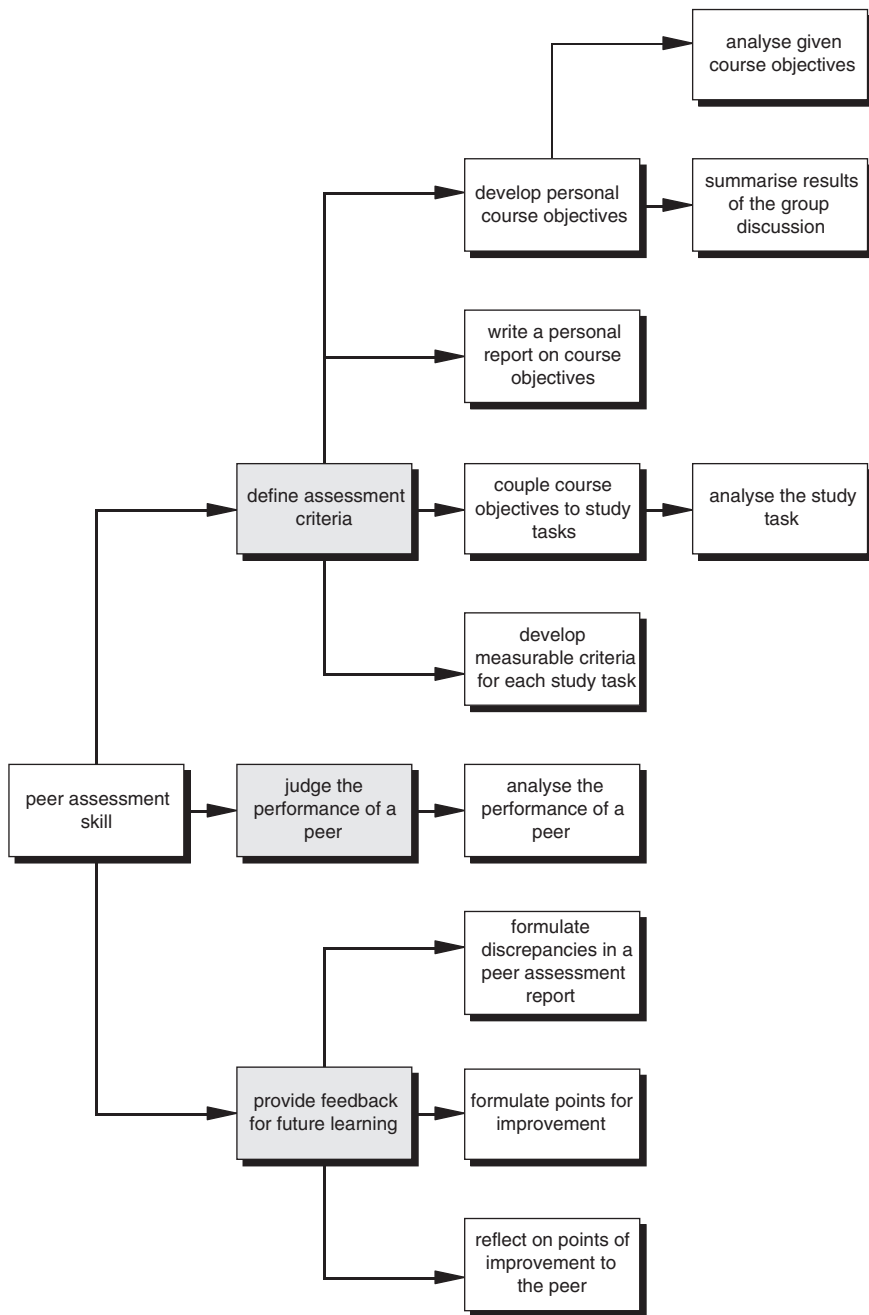
Fig. 12.1 The skills for peer assessment

Figure 12.2 gives an example of authentic assessment about 'using active teaching methods in your teaching to stimulate participation of pupils'.

Research shows that activities as presented in Fig. 12.2 positively affect the development of professional competences as well as the development of the skills to assess peers. For example, discussing assessment criteria about 'using active teaching methods in the classroom' with peers has a positive impact on the skill to use these criteria to assess the performance of peers, but it also improves the student teachers' own performance in the use of active teaching methods in the classroom. Thus, training in peer assessment skills improves at least two skills: the skill to assess work of peers and a domain-specific skill.

Different Modes of Peer Assessment

Depending on the goals of the curriculum, different forms of peer assessments can be introduced. Four considerations are central in deciding what kind of peer assessment should be used.

The first consideration concerns the decision for assessing products or processes of peers. When students assess their peers, the object of assessment is a certain product or a process. A process-oriented peer assessment is useful when the free-rider effect – students who do not participate well in groups – occurs. In process assessment, students can evaluate the contribution of their peers to the collaborative process. However, teacher educators should not use peer assessment as a tool for sanctioning, but as a tool for learning. Discussing assessment criteria with student teachers, focuses them on group roles and group functioning and require students to use the same criteria several times during the collaboration process instead of only at the end, might make the peer assessment more of a learning experience. Also for product-oriented peer assessment, negotiating clear criteria and using multiple peer assessment during one course will improve the positive effects of peer assessment.

A second consideration is whether the peer assessment should be qualitative or quantitative. Quantitative nominations, rankings and ratings, which mostly only contain a mark without additional feedback, have been found to create quite strong adverse reactions (Hanrahan & Isaacs, 2001). For learning, it is more beneficial to choose a qualitative approach. In line with the peer assessment model, students write an assessment report or orally give their feedback to the peer instead of just giving marks that contain little information. Subsequently, the teacher educator could assess the quality of the peer feedback of each student. Thus, qualitative peer assessment seems more valuable for student learning and development of the peer assessment skills than quantitative peer assessment.

Thirdly, peer assessment can be communicated orally or in writing. Research shows that student teachers initially find it difficult to express their feedback in writing (Sluijsmans et al., 2004). However, reporting feedback face to face in a group also entails insecure feelings. The advantage of an oral assessment is the interaction with the peers so that the evaluation of the performance is a joint product

First level	Description
Define assessment criteria	The students actively participate in a group discussion to reach a common understanding about the assessment criteria for a lesson in which the appropriate active teaching methods are applied.
Judge the performance of a peer	The students individually assess a video of a peer by first analysing the lesson (= performance) and then formulating the discrepancies between the observation and the predefined criteria. The formulated discrepancies are written down in a peer assessment report or orally expressed to the peer.
Provide (anonymous) feedback for future learning	The students write a feedback report that provides feedback for future learning. This feedback: - Confirms that the peer's understanding of what was required in the observed performance was correct. - Helps students to add information to their own knowledge when they experience an information gap. - Encourages the peer to replace the erroneous information with more accurate information.

Second level	Description
Develop 'personal' objectives on the basis of given objectives and group discussion	The students present their personal interpretations of what active teaching methods are and in which class situation these are appropriate and share this with the peers in a group session.
Describe a personal report on the objectives	The students individually write a report that reflects their interpretation of the objectives related to active teaching methods.
Relate objectives to assessment tasks	In collaboration with their peers, the students relate the defined objectives to the lessons they have to carry out to reach the objectives and formulate which part of the lesson contributes to certain course objective.
Develop measurable criteria for each assessment task	In collaboration with their peers, the students list the criteria for the performance assessment in which the students have to demonstrate their understanding and use of active teaching methods in the classroom.
Analyse the performance of a peer	The students individually apply the assessment criteria to the product of a peer and they mark pieces of evidence, within this product, that match an assessment criterion.
Formulate discrepancies in a peer assessment report	The students write an assessment report about the quality of the performance in the classroom which reflects evidence for reaching the desired criteria at a certain level.
Formulate points for improvement	The students write individually a number of points for improvement based on the assessment criteria and the group discussions in which the assessment criteria were decided.
Reflect on points of improvement for the peer	Based on the assessed performance, the students individually present ideas for improvement to the peer.

Third level	Description
Analyse given objectives	The students interpret given course objectives based on prior knowledge and personal values by asking questions as 'What do I know about active teaching methods?' and 'What are my experiences with these methods?'
Summarise results of the group discussion	The students take an active role in the group discussion and write a report which represents the outcomes of the discussion.
Analyse the assessment task	The students discuss the upcoming performance assessment with the peers and formulate common criteria that they have to meet to carry out the performance assessment in a proper way.

Fig. 12.2 Description of the constituent peer assessment skills in the context of 'active teaching methods'

of the student and the peers. This is particularly stimulated in peer assessment tasks in which students discuss criteria and feedback rules.

A final consideration is whether the peer assessment should be anonymous or not. In teacher education, it is important to shift from an anonymous to a non-anonymous peer assessment, because teachers who work together have to learn from each other (Verloop & Wubbels, 2000). Student teachers have to get used to an open discussion about criteria and to giving constructive feedback face-to-face. On the other hand, non-anonymous peer assessment might affect the objectivity of the assessment and hinder students from being completely honest, give good argumentations and develop their feedback skills. In teacher education, a non-anonymous peer assessment seems desirable, but novices in peer assessment benefit more from an anonymous setting, until they are used to giving and receiving feedback.

Guidelines for Peer Assessment

For those teacher educators who want to implement peer assessment, we present some practical guidelines for implementing peer assessment:

- Develop student teachers, skills in peer assessment.
- Determine performance criteria before the course.
- Think backwards starting with analysing the desired performance to define assessment criteria, instead of starting with instructional aims and treating the assessment as an afterthought.
- Stimulate collaborative learning and discussion among peers.
- Create interdependency in which peers feel shared responsibility for the group product, process or assessment.
- Start training students' peer assessment skills in their first year of teacher education.

- Prevent illegitimate use of peer assessment, for example, as a sanction tool or to decrease teacher load.
- Provide training for teacher educators in topics like performance assessment, peer assessment and instructional design to increase the successful use of new assessments.
- Create favourable conditions for implementation, both top-down and bottom-up.
- Aim at using peer assessment also as a tool for summative (certifying) assessment.
- Support students to include peer assessments in their portfolio.

Challenge III: Towards Quality in Assessment

During the transition from a testing culture to an assessment culture, ideas of what constitutes 'good' assessment have changed. The use of new modes of assessment requires teacher educators to re-think how the quality of such assessments should be determined and it is argued that a new system for evaluating assessment quality is needed.

Assessment Programmes

We want to start this section with a word of warning. During the transition from a test culture to an assessment culture, a large number of new and different assessment methods have emerged, a few of which were discussed in this chapter. However, it is unwise to assume that new modes of assessment are the panacea for all assessment problems (Maclellan, 2004). New modes of assessment have their problems, too, and some authors note that the claim that new modes of assessment are more valid and suitable still needs empirical confirmation (e.g. Glaser & Silver, 1994). Moreover, it would be unwise to ignore all knowledge of and experience with traditional tests. Instead, we have to combine the virtues of both traditional tests and new modes of assessment. Competences are such complex entities of knowledge, skills and attitudes that it is often argued that one single assessment method cannot adequately cover all aspects of a competence and a mix of methods should be used instead (e.g. Van der Vleuten & Schuwirth, 2005). Therefore, the use of programmes of assessment, instead of single methods, may be a valuable approach in the assessment culture. Within such a programme, new modes of assessment (both formative and summative assessments) can be combined with more traditional knowledge tests. As such, new modes of assessment and traditional tests are not viewed as alternatives to each other, but they rather play complementary roles (Baartman, Bastiaens, Kirschner, & Vleuten, 2006).

New Assessments, New Quality Criteria?

The quality of traditional tests is generally determined by quality criteria such as validity and reliability, but the question arises as to whether these criteria are sufficient

for new modes of assessment. Because new and other modes of assessment are added to the already existing ones, it is critical to also expand the quality criteria we use to judge the adequacy of these different assessments.

Previous studies have described some problems with regard to the use of reliability and validity to evaluate new modes of assessment (Baartman et al., 2007a). Reliability is the degree to which the same results are obtained at a different time, in a different context or by a different assessor. From a traditional point of view, reliability is defined as test–retest accuracy or inter-rater reliability and is often achieved by standardisation. New assessments are not standardised and do not have one correct answer, which makes 'objective' assessment impossible. Moreover, assessment is used to demonstrate development instead of measuring a stable trait. In new modes of assessment, reliability has to be defined in a different way. Reliable assessments require multiple occasions, multiple contexts, multiple methods and multiple assessors.

When it comes to validity, the main problem is that many different definitions of validity exist, like internal validity, construct validity, face validity. The breadth and complexity of the concept make it difficult to work with it in practice (Crooks & Kane, 1996). A new system for assessment quality should clarify and further operationalise the concept of validity for practical use.

Twelve Quality Criteria for New Modes of Assessment

Based on literature research, an expert consultation and a consultation of teachers (Baartman et al., 2006; 2007a; 2007b), 12 quality criteria for competence assessment programmes were formulated. These quality criteria are put together in a wheel, called the wheel of competence assessment (see Fig. 12.3).

In the wheel of competence assessment, the quality criteria are displayed in circles. Fitness for purpose forms the basic quality criterion for all assessments and is related to the previously mentioned importance of congruence between learning, teaching and assessment (Biggs, 1996). The quality criteria in the inner layer are the more basic quality criteria for all modes of assessment. Comparability and reproducibility are derived from the traditional notion of reliability, but they are defined in a different way. The outer layer of the wheel represents the new quality criteria originating in the new assessment culture. The wheel itself is placed in a broader educational context including the criteria of costs and efficiency and educational consequences, which represent the connection of assessment with other aspects of education as a whole.

1. Fitness for purpose relates to the congruence of learning, teaching and assessment as discussed previously in this chapter. It is crucial that the assessment goals and the assessment methods are consistent with the instructional goals and competences strived for.
2. Comparability addresses the fact that assessments should be conducted in a consistent and responsible way for all students. Even though new modes of assessment are less standardised and may differ between students, key features
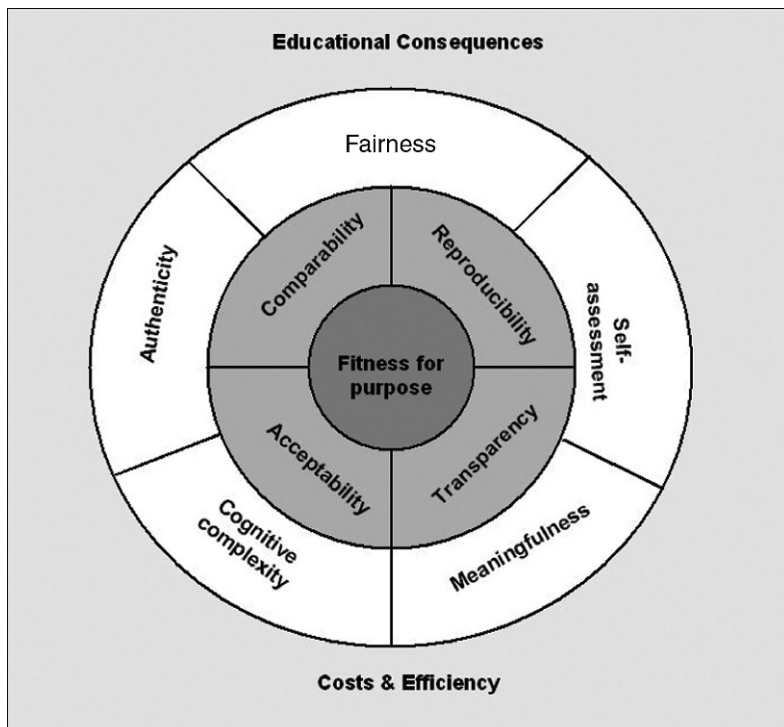
Fig. 12.3 The wheel of competence assessment

have to be consistent for all students. In addition, assessment conditions, procedures and criteria have to be similar for all students.

3. Reproducibility of decisions means that decisions made about students' competence should not depend on coincidence but be based on multiple assessments, multiple assessors and multiple occasions. In new modes of assessment, multiple assessors judge the performance of learners and reach their final decision in an open discussion. It is useful to use assessors with different backgrounds, as these assessors perceive a student's competences from a different point of view and together they can give a balanced decision about the competences of the student.

4. Acceptability means that teacher educators, student teachers and school principals, as future employers, should approve the assessment criteria and the way the assessments are carried out. They have to have confidence in the quality of the assessment methods used. A possible way to increase the acceptability of an assessment is to involve students in the development process of, for example, the assessment criteria, by asking them what they think constitutes a good performance.

5. Transparency relates to the fact that assessments must be clear and understandable to all participants. Assessors and students need to know and understand the

assessment procedures and criteria to be able to prepare for the role of assessor or to adjust their learning process.

6. Fairness means that all students get a fair and equal chance to demonstrate their competences. All students need to be given the opportunity to demonstrate their abilities and to maximise their potential. Therefore, a variety of methods should be used to address the various learning styles or cultural backgrounds of students.

7. Fitness for self-assessment means that assessment should stimulate the development of self-regulated learning skills. Assessment can contribute to these skills by using forms of self-assessment and peer assessment and by allowing students to formulate new learning goals based on their own assessment results.

8. Meaningfulness implies that assessment should have a significant value for educators, student teachers and school principals as future employers and provide a challenging educational experience. Possibilities to increase meaningfulness for students is to involve them in the assessment process (e.g. as described in the section on peer assessment), to let them adjust the assessment to their own personal interests (McDowell, 1995) or to give them an opportunity to decide when they are ready to take the assessment.

9. Authenticity as a quality criterion for assessment has already been addressed earlier in this chapter. It is generally described as the degree in which the assessment reflects the competences needed in the future workplace.

10. Cognitive complexity resembles authenticity in that it also relates to the future professional life, but it focuses on the fact that assessments should elicit the cognitions of practitioners, in this case teachers, to solve problems related to working in education. The use of performance assessments, however, is no guarantee that higher cognitive skills are being measured. To gain insight into the thinking processes of students, Maclellan (2004) suggests to encourage students to clarify the rationale for their answer or action chosen.

11. Educational consequences pertain to the effects the assessments have on learning and instruction (Dierick & Dochy, 2001). A collection of evidence is needed about the intended and unintended effects of the assessment on how teachers and learners adjust their teaching and learning based on their expectations of the assessment. For summative purposes, unintended factors and adverse impact are especially important.

12. Costs and efficiency as a quality criterion is especially important when innovating towards new, more complex assessments. Assessment choices are not only influenced by educational, didactical factors, but also by financial, managerial and institutional factors. This criterion relates to the time and resources needed to carry out the assessment, compared to its benefits.

What Do These Quality Criteria Mean for Teacher Education?

An important new insight in quality evaluation of new assessment is that quality is to a great extent determined by how teacher educators use the assessment instruments.

This implies that teacher educators should be able to critically review (their own) assessment practices and teach about new assessment practices. This can be done by, for example, using a self-evaluation procedure in which all those involved in assessing student teaches, like teacher educators and mentors and school principals, evaluate their own assessments. Self-evaluation seems to improve a critical attitude towards one's own practice, which is crucial for ongoing change and improvement of high-quality assessment in teacher education (see Baartman et al., 2007b).

Conclusion

This chapter discussed three main challenges teacher educators face when developing high-quality assessments in the assessment culture. The goal of this chapter was to encourage teacher educators to reflect on their current or intended assessment practices. We tried to offer practical guidelines to develop more authentic assessments for assessing learning in teaching practice as well as in the teacher education institute, to get student teachers involved in the assessment process through using different kinds of peer assessments and to critically evaluate and improve the quality of assessment programmes. However, assuring authenticity, student involvement and quality is easier said than done. It will require professional development opportunities for teacher educators to enhance their assessment competences. Only then, will they be able to teach as they preach.

References

Baartman, L. K. J., Bastiaens, T. J., Kirschner, P. A., & Vleuten, C. P. M., van der (2006). The wheel of competency assessment. Presenting quality criteria for competency assessment programs. *Studies in Educational Evaluation*, *32*, 153–177.

Baartman, L. K. J., Bastiaens, T. J., Kirschner, P. A., & Vleuten, C. P. M., van der (2007a). Teachers' opinions on quality criteria for competency assessment programmes. *Teaching and Teacher Education*, *23*, 857–867.

Baartman, L. K. J., Prins, F. J., Kirschner, P. A., & Vleuten, C. P. M., van der (2007b). Determining the quality of competence assessment programmes: A self-evaluation procedure. *Studies in Educational Evaluation*, *33*, 258–281.

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, *32*, 347–364.

Birenbaum, M. (1996). Assessment 2000: Towards a pluralistic approach to assessment. In M. Birenbaum & F. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes and prior knowledge* (pp. 3–29). Boston, MA: Kluwer Academic Press.

Birenbaum, M. (2003). New insights into learning and teaching and the implications for assessment. In M. Segers, F. J. R. C. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards* (pp. 13–37). Dordrecht: Kluwer.

Boud, D. (1995). *Enhancing learning through self-assessment*. London: Kogan Page.

Cooper, B. (1994). Authentic testing in mathematics? The boundary between everyday and mathematical knowledge in national curriculum testing in English schools. *Assessment in Education: Principles, Policy & Practice*, *1*, 143–166.

Crooks, T. J., & Kane, M. T. (1996). Threats to the valid use of assessments. *Assessment in Education: Principles, Policy & Practice*, *3*, 265–285.

Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment in teaching in context. *Teaching and Teacher Education*, *16*, 523–545.

Dierick, S., & Dochy, F. (2001). New lines in edumetrics: New forms of assessment lead to new assessment criteria. *Studies in Educational Evaluation, 27*, 307–329.

Elwood, J., & Klenowski, V. (2002). Creating communities of share practice: The challenge of assessment use in learning and teaching. *Assessment and Evaluation in Higher Education*, *27*, 243–251.

Freeman, M. (1995). Peer assessment by groups of group work. *Assessment and Evaluation in Higher Education*, *20*, 289–300.

Gibbs, G. (1992). *Improving the quality of student learning*. Bristol: Technical and Educational Services.

Glaser, R., & Silver, E. (1994). Assessment, testing and instruction: Retrospect and prospect. *Review of Research in Education*, *20*, 393–419.

Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, *52*(3), 67–85.

Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2006). Authentic assessment, student and teacher perceptions: the practical value of the five-dimensional framework. *Journal of Vocational Education and Training*, *58*, 337–357.

Hanrahan, S., & Isaacs, G. (2001). Assessing self- and peer assessment: The students' views. *Higher Education Research and Development*, *20*, 53–70.

Järvinen, A., & Kohonen, V. (1995). Promoting professional development in higher education through portfolio assessment. *Assessment and Evaluation in Higher Education*, *20*, 25–36.

Klenowski, V. (2002). *Developing portfolios for learning and assessment: Processes and principles*. London: Routledge Falmer.

Maclellan, E. (2004). Initial knowledge states about assessment: Novice teachers' conceptualisations. *Teaching and Teacher Education*, *20*, 523–535.

McDowell, L. (1995). The impact of innovative assessment on student learning. *Education and Training International*, *32*, 302–313.

Schön, D. (1987). *Educating the reflective practitioner*. San Francisco, CA: Jossey-Bass.

Sluijsmans, D. M. A., Brand-Gruwel, S., Merriënboer, J., van, & Martens, R. (2004). Training teachers in peer assessment skills: Effects on performance and perceptions. *Innovations in Education and Teaching International*, *41*, 59–78.

Sluijsmans, D. M. A., Moerkerke, G., Dochy, F., & Merriënboer, J., van (2001). Peer assessment in problem based learning. *Studies in Educational Evaluation*, *27*, 153–173.

Verloop, N., & Wubbels, Th. (2000). Some major developments in teacher education in the Netherlands and their relationship with international trends. In G. M. Willems, J. H. J. Stakenborg, & W. Veugelers (Eds.), *Trends in teacher education* (pp. 19–32). Leuven-Apeldoorn: Garant.

Vleuten, C., van der, & Schuwirth, L. (2005). Assessing professional competence: From methods to programmes. *Medical Education*, *39*, 309–317.