Summary

Chapter 1
Over the last decade the landscape of Dutch postgraduate medical training has changed dramatically. It has transformed from loosely-structured master-apprentice style training towards highly structured competency-based, outcome-directed training. One of the most salient features of this transformation is the increase in the number and variety of formative assessment moments: assessments that are aimed to provide feedback that can direct and stimulate learning. In this chapter we explain why the long cherished master-apprentice style of learning had to give way and why formative assessment has attained such a prominent role in Dutch postgraduate medical training.

However, even though the theoretical background for introducing numerous formative assessment moments appeared reasonable and sound, at the introduction little was known about the utility of formative assessment in postgraduate medical education and how assessment can be best positioned when the aim is to optimise learning and training. For this reason we embarked on this research journey, which focuses on the following question:

*How can assessment be deployed to optimise postgraduate medical training?*

Chapter 2
The role of knowledge and the way knowledge should be assessed in postgraduate medical education is subject to discussion. Recent insights from cognitive psychology and the study of deliberate practice recognise that for expert problem solving, a well organised knowledge database is required. Knowledge is usually assessed by modular or final exams, however, these can have unwanted side effects, such as superficial, exam directed study behaviour and the inability to extrapolate test results to the maintained knowledge level over time. For these reasons, longitudinal assessment, like formative knowledge progress testing aimed at steering and directing learning throughout the whole training program, seems a promising approach towards postgraduate knowledge assessment.

Postgraduate obstetrics and gynaecology training in the Netherlands has known a formative knowledge progress test since 1998. For the first study data of 10 years of progress testing were analysed on reliability with Cronbach’s alpha and on construct validity using one-way ANOVA with a post-hoc Scheffé’s test. Average reliability with true-false questions was 0.50, which is moderate at best. After the introduction of multiple-choice questions average reliability improved to 0.65. Construct validity or discriminative power could only be demonstrated with some certainty between training years 1 and 2 and higher training years.

We concluded that both reliability and construct validity of current formative knowledge progress testing in postgraduate obstetrics and gynaecology training is unsatisfactory. Suggestions for improvement of both test construct and test content are provided.
Chapter 3
After evaluating the reliability and construct validity aspect of formative postgraduate knowledge progress testing, we progressed to complement our analysis by evaluating the acceptability and educational impact of the test. Acceptability can be defined as the degree of which the test meets the needs of stakeholders (trainees and educational supervisors) and whether stakeholders endorse the interpretation and consequences of test scores. Educational impact, on the other hand, is the influence of feedback or test scores on directing and structuring study activities. Especially for formative assessment acceptability and educational impact may be at least as important as psychometric test characteristics. For this purpose a questionnaire was developed that targeted both educational supervisors and postgraduate trainees, containing questions on general acceptability, educational impact and acceptability of test content. Ninety percent of trainees and 84 percent of educational supervisors completed the questionnaire.

The general acceptability of formatively used progress testing is good; however, the self-reported educational impact is limited. Furthermore, trainees query the validity of test content. We discuss the importance of feedback quality and the effect of grading. Furthermore we start a debate on whether, for a genuine effect on learning, formative assessment should have consequences, either by entwining the assessment firmly with the training program or by linking the assessment to a summative standard.

Chapter 4
After establishing that formative postgraduate knowledge progress testing has a barely acceptable reliability and a limited construct validity, we sought ways to improve the test. Trainees objected to the number of factual and percentage questions in the test. With the availability of fast internet on the work-floor they claimed that, if needed, they were able to look information up with two ‘mouse-clicks’. One way to improve construct validity and reliability of a test is to improve the authenticity of a test, and so we hypothesised that allowing internet access for a limited amount of time during the progress test would improve the perception of authenticity (face validity) of the test, which would in turn improve the construct validity and reliability of postgraduate progress testing.

Postgraduate trainees taking the yearly knowledge progress test were asked to participate in a study where they could access the internet for 30 minutes at the end of a traditional pen and paper test. Before and after the test they were asked to complete a short questionnaire regarding the face validity of the test. Mean test scores increased significantly for all training years. Trainees indicated that the face validity of the test improved with internet access and that they would like to continue to have internet access during future progress testing. However, internet access did not improve the construct validity or reliability of the test.

Chapter 5
Recent changes in postgraduate medical training curricula usually encompass a shift towards more formative assessment, or assessment for learning. However, even though theoretically well suited to postgraduate training, evidence is emerging that engaging in formative assessment in daily clinical practice is complex. For this reason, we set out to
explore trainees’ and supervisors’ perceptions regarding what factors determine active engagement with formative assessment using focus group interviews.

A total of seven focus group interviews were held, three with educational supervisors and 4 with postgraduate trainees. Three higher-order themes emerged that determine active engagement in formative assessment. First of all: the individual perspective on feedback, which appears to be largely determined by ownership of the trainee and mastery versus performance goal orientation of trainee and/or supervisor. Furthermore, the supportiveness of the learning environment plays an important role, for example dedicated teaching time, motivated supervisors and clear standards and procedures. Last, but not least: the credibility of feedback and/or feedback giver determines whether the trainee will make an effort to act on the feedback. We conclude that engaging in formative assessment with a genuine impact on learning is complex and quite a challenge to both trainees and supervisors. Individual perspectives on feedback, a supportive learning environment and credibility of feedback are all important in this process. Every one of these should be taken in account when the utility of formative assessment in postgraduate medical training is evaluated.

Chapter 6

In contemporary postgraduate medical education competency-based, outcome-focused training has replaced more traditional master-apprentice style training. This change requires a different approach to the assessment of clinical competence, especially given the decisions that must be made about the level of independence allowed to trainees. During the focus groups study into the utility of formative assessment, questions were asked about the role of formal assessment preceding progressive independence during clinical training. These questions yielded so much information that we reported on these in a separate paper.

Two higher-order themes emerged: factors that play a role in determining the level of competence of a trainee in a clinical procedure, and factors that determine the level of independence granted to a trainee or acceptable to a trainee. The level of competence as experienced by trainees was dependent on numbers of performed procedures, awareness of personal limitations and previous feedback. Supervisors determined the level of competence of a trainee by combining direct observation, information from other staff, knowledge assessment, number of performed procedures and the result of the trainees’ self-assessment. Furthermore, both the local context and the procedure itself (complexity, frequency of appearance) were determinants of competence. At the same time it appeared that achieved competence does not automatically translate into more independence. For this trainees need to feel confident and safe and supervisors need to be courageous and able to balance patient safety issues. Furthermore, competence was deemed case-specific and could fluctuate with the time of day (during or after office hours).

Remarkably trainees did not see added value in formal assessment of skills preceding more independence, as they feared to be left on their own to soon. We discuss the implications of our findings for the assessment of clinical competence and argue that society deserves a transparent assessment structure for the assessment of skills, especially preceding progressive independence. Additionally we provide suggestions for a fair and defensible performance assessment structure that pays explicit attention to the preservation of progressive independence.
Chapter 7

Many countries recently revised the assessment structure of postgraduate medical training with the aim to make training more effective, efficient and transparent. However, though theoretically well grounded, at the start of the reforms little information was available on the educational impact of recent assessment program changes. We set out to systematically review the literature for evidence of the educational impact of assessment instruments used in postgraduate medical training. The studies were assessed using best evidence in medical education (BEME) quality indicators and Kirkpatrick’s evaluation model.

Forty seven articles were retrieved, of which 5 were a systematic review that covered components of the current review. The majority (26) of studies reported the perceptions of trainees of the assessment instrument under investigation, half of which reported a perceived positive impact on learning. Sixteen articles were retrieved that addressed the impact of assessment at the level of improved knowledge or skills, behaviour and/or patient outcomes, all of which found a significant positive effect of assessment on learning. Realist synthesis identified several factors associated with achieving a positive effect on learning through assessment: comparison to a gold standard, sufficient allocated time for structured feedback, feedback is part of a feedback cycle.

A defendable postgraduate assessment structure, which can exert a genuine impact on postgraduate medical training, includes the following elements: multisource feedback, direct observations of patient encounters and skills followed by structured feedback, assuming that sufficient time is allocated and that feedback is part of a feedback cycle, formative knowledge assessment and certifying examinations.

Chapter 8

This research project started with the following question: How should assessment be deployed in order to optimise postgraduate medical training? Our research has provided us with the following insights into how a measurable educational impact can be achieved using formative postgraduate assessment:

1. The assessment compares trainee performance to an agreed-upon standard.
2. The feedback that follows assessment is part of a feedback cycle (assessment – feedback – individual learning goal - period of practice – assessment).
3. The feedback cycle is pursued in a sustainable trainee-supervisor relationship.

From these insights we come to the following recommendations for Dutch postgraduate assessment:

• Establish a healthy mix of formative and summative assessments.
• Continue workplace-based assessment but ensure that it is firmly embedded in a system that facilitates its use in a way that can have a genuine impact on postgraduate training (sustainable trainee-supervisor relationships).
• Entrusted professional activities should be preceded by assessment of knowledge, skills and performance.
• Combine knowledge progress testing with summative modular testing or certifying exams.
• Use portfolios for the collection of evidence of competence, but not for the assessment of competence.

Summary