

The impact of programmatic assessment on student learning: theory versus practice

Sylvia Heeneman,^{1,2} Andrea Oudkerk Pool,^{1,3} Lambert W T Schuwirth,^{2,3,4} Cees P M van der Vleuten^{2,3} & Erik W Driessen^{2,3}

CONTEXT It is widely acknowledged that assessment can affect student learning. In recent years, attention has been called to 'programmatic assessment', which is intended to optimise both learning functions and decision functions at the programme level of assessment, rather than according to individual methods of assessment. Although the concept is attractive, little research into its intended effects on students and their learning has been conducted.

OBJECTIVES This study investigated the elements of programmatic assessment that students perceived as supporting or inhibiting learning, and the factors that influenced the active construction of their learning.

METHODS The study was conducted in a graduate-entry medical school that implemented programmatic assessment. Thus, all assessment information, feedback and reflective activities were combined into a comprehensive, holistic programme of assessment. We used a qualitative approach and interviewed students ($n = 17$) in the pre-clinical phase of the programme about their perceptions of programmatic assessment and learning approaches. Data were scrutinised using theory-based thematic analysis.

RESULTS Elements from the comprehensive programme of assessment, such as feedback, portfolios, assessments and assignments, were found to have both supporting and inhibiting effects on learning. These supporting and inhibiting elements influenced students' construction of learning. Findings showed that: (i) students perceived formative assessment as summative; (ii) programmatic assessment was an important trigger for learning, and (iii) the portfolio's reflective activities were appreciated for their generation of knowledge, the lessons drawn from feedback, and the opportunities for follow-up. Some students, however, were less appreciative of reflective activities. For these students, the elements perceived as inhibiting seemed to dominate the learning response.

CONCLUSIONS The active participation of learners in their own learning is possible when learning is supported by programmatic assessment. Certain features of the comprehensive programme of assessment were found to influence student learning, and this influence can either support or inhibit students' learning responses.

Medical Education 2015; 49: 487–498
doi: 10.1111/medu.12645

Discuss ideas arising from the article at
www.mededuc.com/discuss.



¹Department of Pathology, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands

²Faculty of Health, Medicine and Life Sciences, School of Health Professions Education, Maastricht University, Maastricht, The Netherlands

³Department of Educational Development and Research, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands

⁴Health Professions Education, School of Medicine, Flinders University, Adelaide, South Australia, Australia

Correspondence: Sylvia Heeneman, Department of Pathology, Maastricht University/MUMC, Peter Debye laan 25, 6229 HX Maastricht, The Netherlands. Tel: 00 31 43 387 6629; E-mail: s.heeneman@maastrichtuniversity.nl

INTRODUCTION

Although it is widely acknowledged that assessment has an impact on student learning,^{1–3} the interplay between assessment and student learning is complex⁴ and the mechanisms through which assessment drives student learning, be it positively or negatively, are still highly unpredictable.⁵

Theoretical models have been proposed to help elucidate the effects of assessment and it has become clear that assessment affects learning before, during and after the assessment activity, producing what are described as pre-, pure and post-assessment effects on learning.³ The challenge of any assessment practice is to predict and shape these effects of assessment on learning. In the teaching environment, there has been a shift towards a constructivist perspective in which students are regarded as active acquirers of their own knowledge, skills and competencies. This contrasts starkly with assessment that is still mainly based on the behaviourist notion of testing whether or not a student has learned enough.^{6,7} Therefore, calls have been made for a move from assessment *of* learning, in which 'simple' pass/fail decisions are made, to assessment *for* learning, in which the assessment environment encourages students to feel responsible for and direct their own learning.⁸ Assessment for learning needs to be translated into curricula and concrete educational settings. To render this possible, students should receive enough and meaningful information on their performance. A second *sine qua non* is that students actively abstract meaning from this information and use it to manage their learning and gain ownership of their performance improvement, allowing them to perform better the next time and to progress as learners.⁹ For the transposition of assessment for learning into curricula, a programmatic assessment approach has been suggested.⁸ In this approach, a variety of informative assessment activities are purposefully selected, combined and arranged in time to constitute a comprehensive programme of assessment that provides a longitudinal flow of information about the student. At the same time, the student is required to use this information to self-direct learning and to learn from assessment. Moreover, instead of single assessment activities, the aggregate of all information available is used to come to pass/fail or high-stakes decisions, or to decide on promotion.^{10,11} The programmatic assessment approach has the potential to increase

the robustness of the assessment process itself and can foster the constructivist notion of how learning takes place. Although it is conceptually attractive, little research into the intended effects of a holistic, programmatic assessment approach on students and on how they construct their learning has been conducted.

A number of studies have reported on elements or characteristics of assessment environments that may be conducive to learning. These include a supportive learning environment and credible feedback,¹² formative assessment, narrative feedback and the opportunity to use feedback in reflection.¹³ Case reports of curricula that have implemented programmatic assessment show that formative assessment, narrative feedback and reflection do indeed encourage assessment for learning.^{14,15} Bok *et al.*,¹⁶ however, showed that in the context of programmatic assessment using formative feedback in the clinical veterinary workplace, students still perceived the feedback as summative, which suggests that the intended effects of programmatic assessment were not in keeping with practice. All in all, these studies show the potential of comprehensive assessment programmes to support assessment for learning. Yet, these conclusions must be qualified by the facts that formative procedures were perceived as summative,¹⁷ and the studies were mainly performed in workplace-based assessment settings.^{15,17} The focus on assessment for learning in the workplace-based setting is understandable because the workplace provides ample opportunity for a learner to self-regulate and to shape future learning opportunities. Although in the earlier stages of a (medical) curriculum more conventional educational methods are in place, in this setting it is also important to study the potential of programmatic assessment to foster student learning in line with current constructivist insights. Therefore, we performed a study in a pre-clinical setting to gain more insight into the following research questions: (i) which elements of the comprehensive programme of assessment do students perceive as supporting or as inhibiting their learning? (ii) What are the factors that students consider important for the active construction of their learning in an assessment for learning environment?

To examine these issues, we used a qualitative approach and conducted individual interviews with students to find out about their perceptions of programmatic assessment and their learning approaches.

METHODS

Setting

The research setting is a 4-year, graduate-entry Master's programme (50 students per year), designated the Physician–Clinical Investigator (P-CI) programme, at Maastricht University, the Netherlands. The curriculum is competency-based and uses the CanMEDS framework.¹⁸ In the programme, programmatic assessment is implemented; the structure and components are shown in Table 1.¹¹ The programme of assessment is designed to foster assessment for learning through the following features: (i) diversity in assessment and assignment formats is used to optimise the provision of information and feedback to individual learners about their progress and competency development throughout the year; (ii) narrative feedback from peers and teachers is used to provide rich information on performance; (iii) all feedback and information are combined into a portfolio and used by students for personal reflection, to self-regulate their learning and to follow up on learning through specific learning objectives; (iv) all students are coached by a personal mentor; (v) assessments and assignments or tasks (as shown in Table 1) are not directly summative in the sense that they do not result only in a pass/fail decision for the student, but are meant to be informative and contribute to the eventual summative decision at the end of a study phase, and (vi) a high-stakes decision for promotion to the next phase is based on all information collected and collated in the portfolio. Students are expected to seek additional feedback and show progress if the assessment or feedback information indicates insufficient knowledge on a specific subject or skills. For certain assessment activities, remediation sessions are arranged if the curriculum does not give the student another opportunity to show sufficient knowledge or skills.

Methodology and data collection

A qualitative approach was chosen to explore the students' own experiences in their natural context,¹⁹ using an interpretative, constructivist approach.²⁰ Individual semi-structured interviews with 17 students (seven men and 10 women) were conducted. Students were sampled using maximum variation sampling to ensure that the multiple perspectives of individuals were represented.²¹ Between November and December 2013, the second author (AOP) interviewed individual students in the

pre-clinical Year 2. Year 2 students were selected because they have at least 1 year of experience with programmatic assessment. A total of 19 students were invited; one student did not respond to the invitation and one student responded positively initially, but did not respond to the second e-mail. Thus, 17 students responded positively and were interviewed. The mean age of the respondents was 23.9 years (range: 22–26 years) and the male : female ratio (40 : 60) was similar to this ratio in the programme (Table 2). The interviews lasted 60–90 minutes. The interview questions were semi-structured (Table S1) and were based on fundamental concepts of programmatic assessment, feedback, the effects of assessment on learning, and self-regulation of learning.^{3,4,11–13,15,22–27}

Ethical considerations

Participation was voluntary. Students were assured of confidentiality and asked to sign an informed consent form. The study was approved by the ethical review board of the Dutch Society for Medical Education (approval no. NVMO-ERB-276). The researchers were educationalists (EWD, CPMvdV, LWTS), a psychologist (AOP) and a biologist with an educational background (SH). EWD, CPMvdV and AOP had no direct contact with the students in the programme. However, SH did, as the programme director.

Analysis of interviews

Interviewing and analysis were conducted iteratively in order to facilitate the exploration of new themes and categories in subsequent interviews. Verbatim transcripts of the interviews were made and analysed using a type of theory-based thematic analysis, template analysis, in which a succession of coding templates, consisting of hierarchically structured themes, were applied to the data.²⁸ Template analysis starts with a small set of predefined codes that guide analysis. This set of predefined codes (Table S2) was based on the set of papers that had guided the development of the interview questions (see above). Independent analysis of interviews 1–7 (by SH and AOP) using the predefined codes led to the development of an initial template that was discussed with another researcher (EWD) and then used in interviews 8–14. This led to a final template that was discussed with the research team (SH, AOP, EWD, CPMvdV, LWTS). Theoretical saturation (defined as occurring when [i] new data could be fitted in categories already devised, and [ii] no new insights were

Table 1 Assessment programme for the pre-clinical part of the curriculum (first 1.5 years)

| Learning activities | | |
|---|---|---|
| Type | Format | Assessment and evaluation activities |
| PBL Years 1 and 2 Organ-based blocks Year 1 Care-in-practice block Year 2 Science-in-practice block | Year 1 Paper-based cases Year 2 Patients in teaching hospital as cases | Years 1 and 2: Knowledge test Timing: only end-of-block; mid- and end-of-block; multiple smaller tests (mini-tests) Format: written multiple-choice, open-ended questions; oral examinations with different examiners Information: either grade (0–10) or mark (pass/fail or pass/fail/good); narrative feedback Years 1 and 2: OSCEs Timing: two stations per block Information: narrative feedback and mark (Year 1) or grade (Year 2) Years 1 and 2: Feedback on performance in PBL group Timing: mid- and end-of-block by peers and tutor Format: forms contain open-text fields for the seven CanMEDS competencies narrative, written feedback form (two peers) or (digital, anonymous) multi-source feedback (five peers) |
| Practising medicine assignments (ethics, prevention, etc.) | Group or individual assignments in the blocks | Timing: one or two assignments per block Information: either grade (0–10) or mark (pass/fail or pass/fail/good); narrative feedback Group work: peer feedback |
| Clinical investigator assignments | Group or individual assignments In the blocks and longitudinal | Timing blocks: one or two assignments per block Timing longitudinal: Year 1: four reports on attended research presentations, one review on translational medicine. Year 2: one (semi) meta-analysis Information: either grade (0–10) or mark (pass/fail or pass/fail/good); narrative feedback Group work: peer feedback |
| Progress test | | Timing: four times per year Format: a multiple-choice test that tests the complete knowledge domain ⁴⁵ Information: both individual and group performance at discipline level |
| <p>Portfolio</p> <p>Reflection on information by student</p> <p>Setting and follow-up of learning objectives by student</p> <p>Coaching by a mentor (meeting five times in Year 1, three times in Year 2)</p> <p>Independent feedback halfway through year by a second mentor</p> <p>Portfolio assessment</p> <p>End of year</p> <p>Recommendation on student progress and competency level by the student's mentor</p> <p>Recommendation on student progress competency level by group of mentors</p> <p>Assessment by independent assessment committee, using all the information in the portfolio, triangulating information of mentor recommendations</p> <p>High-stake decision, pass/fail/good for promotion to next year</p> | | |

OSCE = objective structured clinical examination; PBL = problem-based learning.

Table 2 Summary of interviewee characteristics based on gender and portfolio assessment result in Year 1

| Portfolio assessment outcome in Year 1 | Gender | |
|--|---------------------|-------------------------|
| | Male | Female |
| Fail | Interviews 4, 12 | Interviews 2, 6, 11 |
| Pass | Interviews 1, 5, 10 | Interviews 3, 9, 16, 17 |
| Good | Interviews 7, 14 | Interviews 8, 13, 15 |

obtained, no new themes identified and no new issues arose regarding a category of data)²⁹ was reached after interview 12. All interviews were then reread by SH and AOP to ensure that no relevant information had been missed. The final template was confirmed by analysis and coding of interviews 15–17 and is represented in Table 3. After a thorough discussion of the complete dataset among the entire research team, the analysis was advanced from the themes to an interpretation of the relationships between learning responses and information generated by assessment and feedback activities embedded in a comprehensive programme of assessment.

RESULTS

The following paragraphs discuss the findings in light of the research questions. The outcomes revealed a strong correlation between the aspects of programmatic assessment identified by students as productive or counterproductive (research question 1) and the extent to which learning was actively constructed (research question 2). This calls for a brief spelling out of the supporting and inhibiting elements of the comprehensive programme of assessment. The subsequent paragraphs will present the three overarching layers of learning perceived by students as important for the active construction of their learning in an assessment for learning environment.

Elements of the comprehensive programme of assessment considered to support or inhibit learning

Evaluation and assessment activities from the comprehensive programme of assessment, such as feedback, portfolio, assessments and assignments, had

both supporting and inhibiting effects. This also held true for factors inherent in the individual student and in the programme, such as social interaction amongst students and the curriculum. Table 3 summarises these findings. Hence, evaluation and assessment activities can both support and inhibit learning. For instance, the anonymous character of multi-source feedback activities encourages students to give feedback that they would communicate less easily in face-to-face sessions, but may overwhelm the student who is receiving the feedback and cause him or her to be less prone to accept it as the feedback has not been mentioned before. The analysis showed that the supporting and inhibiting elements of the comprehensive programme of assessment influenced the construction of learning, as the next paragraphs will explain in more detail.

Layers of learning considered important for the active construction of learning in an assessment for learning environment

Students perceived several ‘layers of learning’. These included learning activities instigated by upcoming assessments and assignments, learning during the actual assessment, and learning from the information and feedback generated by the comprehensive programme of assessment. Analysis indicated that these aspects of learning were influenced by how students construct and participate in their learning and the (supporting and inhibiting) characteristics of the comprehensive programme of assessment, thus yielding an environment reflecting a combination of assessment *of* and assessment *for* learning.

Learning activities instigated by assessment and assignment tasks: pre-assessment effects

Assessments or assignments were designed to give information, as formative tasks, and to be used in the portfolio. This was perceived differently by most students because these tasks were seen to be associated with summative pass/fail decisions. This resulted in pre-assessment effects of increased study activity as the assessment day drew nearer. What hindered students’ notions of ‘true’ formative assessment were the perceived signals of summative pass/fail decisions in the programme, such as the organisation of remediation sessions, which instigated the perception of having to pass a classical ‘resit’:

I think, in the end, passing and failing are still part of it as far as I can see. And maybe that is

Table 3 Elements of the comprehensive programme of assessment perceived as supporting or inhibiting learning

| | Supporting learning | Inhibiting learning |
|--|---|---|
| Evaluation and assessment activities | | |
| <i>Feedback</i> | | |
| Peers | Supporting competency-development; use for reflection in portfolio through follow-up of information | Social interaction ('getting back'); frequency too high; 'looking for improvement points'; perceived feedback 'duplication' by peers |
| Teacher | Expert view on competency-development; use for reflection in portfolio through follow-up of information | Quality; lack of feedback; timing (too late); inconsistencies between teachers in feedback; unfamiliarity with all competencies |
| Format | Diversity: narrative and oral feedback and multi-source feedback (anonymity helps in the giving of 'difficult' feedback) | Anonymity in multi-source feedback (receiver can experience certain feedback as unexpected, not heard before in narrative/oral face-to-face feedback) |
| <i>Portfolio</i> | | |
| Reflection | Valuable tool for competency development (internal motivation); awareness; follow-up of information (feedback, marks, qualifications); personal approach (students feel autonomous on the topics of reflection) | Time needed (busy programme); frequency too high; predominant external motivation/attribution |
| Assessment | Independent advice of second mentor; motivation to do well overall | Perception of 'writing what is needed for reflection to pass'; perception of student 'standardisation'; perception that certain topics are obligatory for reflection (hindering autonomy); lack of individual decision moments for modules (less motivation to learn); assessment procedure not transparent |
| Mentor | Discussion of reflection; follow-up of information; coaching of personal development; first contact | Inexperience (mentor and student) |
| Format | Structure (keeping overview) | ICT platform (bugs); structure (limited autonomy) |
| <i>Assessment</i> | | |
| Format | Oral examinations (immediate feedback, sense of agency); OSCEs (perception of mastery, relevance for profession); learning from remediation (gaining new insights) | Questioned validity of some formats; opinionated/biased examiner (oral examination); stress for assessment moments; content of assessment not matching expectancy |
| Information (marks/grade, qualification, feedback) | Mark (gives clarity) in combination with feedback; use for follow-up to next assessment moment; valued relevance for future profession | Grade (qualification), especially if pass/fail only (less drive for performance); limited perception of formative assessment |
| Strategy | Deep/sustained learning (oral examinations); continuous learning (mini-examinations); knowing what to expect in assessment | Conflicting deadlines (busy programme) |
| <i>Assignment</i> | | |
| Information (marks/grade, qualification, feedback) | Mark (gives clarity) in combination with feedback; use for follow-up to next assignment | Grade (qualification), especially if pass/fail only (less drive for performance); perception of formative assessment as summative decision points |
| Format or content | Valued autonomy for content | Insufficient instruction; limited autonomy for content; need to prioritise (busy programme) |

Table 3 (Continued)

| Supporting learning | | Inhibiting learning |
|--------------------------------------|--|--|
| Purpose | Valued relevance for future profession | Lack of embedding in programme |
| <i>Student and programme factors</i> | | |
| Personal/student | | |
| Student interaction | Learning from/with peers | Friendship interfering with (honest) feedback |
| Personal Motivation | Individual progress (competencies/'self') For future profession | Stress (lack of time) Focus on envisioned future discipline early on (less effort for other disciplines) |
| Programme | | |
| Teaching environment | PBL; learning in teaching hospital; expert tutors (most of the time) | No direct observation (some PBL sessions, teaching hospital) |
| Curriculum | Combination of physician and clinical investigator element | Programme very busy (need to prioritise); relevance of certain course elements; too little priority given to knowledge acquisition |

OSCE = objective structured clinical examination; PBL = problem-based learning.

what it comes down to, in the sense that if you do not pass your knowledge tests, you have to re-sit them. So, the concept of pass or fail still exists really. (Interview 11)

Incentives for learning differed amongst students and were often explained by students as ‘wanting to know the medical knowledge for your own good’ and motivation for succeeding in the profession in the future. Motivation to achieve a grade sufficient to allow a ‘pass’ was linked with having a busy programme, lack of available time or competing deadlines. Students associated grading as pass/fail only with loss of information. The message that a ‘pass’ was enough did not motivate them to strive for excellence. This tendency to aim for acceptable standards rather than excellence was especially dominant when the assessment task had to compete with other (assessment) activities at the same moment, as indicated by this student:

With the in-block assessments it’s always up to standard or not up to standard, and I just know that I scored up to standard quite easily, so it does influence me, in the sense that if I had some other assignment simultaneously with a test I would think: oh well, getting up to standard isn’t too hard, so I won’t go for a high grade, so then I can spend less time on it. (Interview 4)

Learning during assessment activities: pure assessment effects

All students felt that oral examinations were very valuable for learning. These oral examinations also referred to other competencies, such as those associated with self-presentation and communication. The expectancy that one would be required to explain mechanisms instead of facts, and in front of an expert, constructed the students’ learning towards a deeper understanding. In addition, a learning effect of the oral examination itself was noted, as the interaction with the expert had a direct learning effect (i.e. a pure assessment effect):

If I don’t understand something during an oral, I can ask questions straight away like: but what do you mean by this or how exactly does that work? ...it’s more that you get more feedback like: OK, this isn’t quite right, this works like this, and that makes you think: all right, OK, or I did it wrong, or yes of course you’re right, but then you know where you stand. (Interview 16)

Learning from information and feedback generated by the comprehensive programme of assessment: post-assessment effects

Peer feedback was valued as a good source of material for reflection, and was noted for usefulness for

competencies related to group work, such as professionalism and collaboration. Teacher feedback was valued differently by some students because of the credibility of the teacher's expert opinion and also the content of the feedback. Students mentioned several aspects of implementation that limited the effectiveness of feedback. Feedback was too densely planned in the programme, leading to a weariness of delivering and receiving feedback and reflecting on the feedback. In addition, the density of feedback affected its quality; sometimes the same feedback was repeated as peers reproduced issues that had already been raised in feedback rounds, such as in the problem-based learning (PBL) groups ('copying' of feedback). In addition, students felt pressurised to offer feedback, which, as they said, led to feedback on details ('searching for improvement points'), not all of which were effective for learning:

I just find that there comes a point where you feel like: I still have to bring up some points for improvement, and then you have to dig deep to find some little detail [...] And then it's very tricky to determine if the comment concerns a genuine issue or if it's just far-fetched. (Interview 10)

The aggregation of all assessments, assignments and feedback information in the portfolio, and its use in decisions on promotion, did not affect learning, as indicated by this student:

I don't even think about the grades really. Somehow or other. I don't know, I just think, like yeah, because that's the way it is really, that the portfolio consists of the different parts of the programme. So, in the end you get grades for the programme parts, but then through the portfolio. I consider that as more or less the same thing. (Interview 4)

Students indicated that reflection was driven by both external and internal motivations. The notion that one had to use all feedback and make learning objectives even if the subject was not considered as a 'big issue', and the high frequency of feedback and reflection activities, were perceived as external motivators for the reflective work in the portfolio. The portfolio was perceived as labour-intensive and students noted a negative interaction between it and the demands of a busy programme. This led to an appraisal of working on the portfolio, in which work done was weighted according to the work needed to pass the portfolio assessment. This was associated with a perception of decreased or limited intrinsic motivation for working on the portfolio. In

addition, some competencies, such as that of 'health advocate', were less recognised in the curriculum, and this also led to reflection driven by external motivation. A small subgroup of (male) students indicated that the portfolio as a whole was a negative experience, driven mainly by external motivation. This was linked to the perceived compulsory nature of the process, a tendency to use feedback only when it was valued as credible by the student and to ignore feedback that was not credible, and the use of external attribution (e.g. the quality of the assessment was seen as being responsible for the poor results) to explain deficiencies or learning needs. It seemed that elements from the programme of assessment perceived as inhibiting dominated the learning response:

The difference in marks between those four reports did not align with my own perceptions, how much time I spent on it, how good I thought they were, so that was for me like: yes well, these assessors used double standards, and this was just very annoying. So I wrote this in my portfolio and then all of a sudden there were two issues for which I used external factors to explain what happened. (Interview 4)

All students reported benefiting from meetings with their mentors and having learned from the discussions and feedback. Most students perceived the reflective writing for the portfolio as valuable and as useful for learning from the feedback and for the self-regulation of learning:

Often you think that without the portfolio it would be easier to think: OK good, I've gotten this feedback, fine, we'll tuck it away there. But now you are challenged to link feedback, look for common threads but also for conflicting points and to think about them. I think it's useful, particularly for this purpose. (Interview 8)

DISCUSSION

In this study, we explored the factors that students consider important for the active construction of their learning when assessment, assignments and feedback are embedded in a comprehensive programme of assessment. The study, moreover, relied on the theoretical assumption that such a programmatic assessment approach would generate a continuous flow of information that the student could use to self-direct learning and to learn from assessment.^{9–11} In response to research question 1,

we identified elements that supported or inhibited learning which related not only to the assessment programme itself, such as assessment format, feedback from peers, reflection and guidance by a mentor, but also to the individual student or to the programme, such as student interaction and the curriculum. The analysis revealed that whether or not an element would support or inhibit learning depended not as much on its nature as it did on how it was perceived by the individual learner. When the learner understood its purpose, he or she would buy into it and, consequently, the element would become meaningful to learning. This attribution of relevance by the learner correlated positively with the extent to which learning was actively constructed (research question 2). Construction of learning was shown to vary in accordance with whether assessment and assignments were mainly viewed as summative, and whether learning for assessments was determined by behaviourist types of pre-assessment effect. Thus, findings showed that: (i) oral examinations as part of programmatic assessment had positive, pure assessment learning effects; (ii) the grading system (i.e. pass/fail only or marks) was an important determinant of the learning response; (iii) motivation for learning and reflective activities was extrinsic for some students, and (iv) reflective activities in a portfolio were regarded positively and were seen as supportive of the follow-up of the information and feedback generated by the comprehensive programme of assessment, evoking a post-assessment learning effect and self-regulation of learning.

It was evident that the formative nature of the assessment and assignment activities was not regarded as such by the students; rather, these were interpreted as summative assessment activities for which pre-assessment effects prevailed.⁴ This is in line with the findings of other studies¹⁶ and has also been observed for other assessment tools intended to give feedback and to serve as types of formative assessment, such as the mini-clinical evaluation exercise (mini-CEX). The mini-CEX has been in use for several years and is based on a large body of educational evidence, yet trainees have difficulties in understanding and exploiting its formative purpose and regard it as a tick-box exercise. Previous research has shown that both during and after its implementation, both residents and assessors require ongoing training to sustain understanding of the informative purpose of the mini-CEX and the importance of providing immediate high-quality feedback.³⁰

It is not uncommon for the translation of theoretical concepts into practice not to go as planned. Studies on the implementation of PBL,³¹ competency-based learning,^{16,32} postgraduate training³³ and workplace-based assessment³⁴ have shown that a number of factors, such as sufficient student and faculty member training, buy-in from stakeholders, the perceptions of students and teachers of theoretical concepts, affinity with the educational reform, and the availability of time and money, are common components in the success of educational reforms. In the current study, these implementation-related factors were recognised in addition to the effects of procedures and a busy programme on access to time or willingness to provide feedback, lack of affinity or negative motivation, and a persistent message of summative assessment (e.g. as a result of the organisation of remediation sessions). It is very clear that the programmatic assessment model requires careful implementation.

Compounding the potential impacts of these implementation-related factors, the concept of formative assessment may also have played a role in the persistent perception of formative tasks as summative. The theoretical concept of formative assessment is not undisputed and denominating it as an 'assessment' may interfere with the actual perceptions of both students and teachers and its impact on learning.^{35,36} It should not be forgotten that these students have been immersed in a classical, summative pass/fail system during their entire previous education. Their teachers in the programme still insist that they pass examinations with a certain grade or attend remediation sessions, rather than letting students regulate their own learning and follow-up on learning objectives. It may not be easy to break away from conventional assessment-informed practices and move towards a vision of assessment as a valuable information-provider. Taras^{36,37} argued that all assessment starts with summative assessment, in which a judgement is given, and that formative assessment actually represents the subsequent feedback loop, in which the summative assessment and feedback are used by the student to monitor the desired and actual performance, and make learning objectives for follow-up. This mirrors what was seen in this study; the end-of-block assessments and certain assignments were clearly perceived as providing for summative judgements, but the information and feedback derived from these assessments were used to identify learning needs, follow up on feedback and ask for new feedback. This applied to feedback from both peers and teachers. This effect could be termed the 'formative assessment loop', as

inspired by Taras,^{36,37} or as a post-assessment effect, whereby learning takes place after the assessment task has been completed or feedback received.³ This has also been described as the catalytic effect of assessment, in which the assessment generates feedback, which, in turn, creates the opportunity to enhance and support education, and to move learning forward.³⁸ In summary, designating assessment and assignment tasks as 'formative' may not align with the theory, and will not match the perception of students, but using a programmatic assessment approach may still lead to learning via a post-assessment effect. Further research should focus on the question of whether and in which conditions a low-stakes formative assessment is indeed perceived as a low-stakes assessment. The reflective activities in the portfolio represented an essential element for the post-assessment effect on learning, as has been shown in other studies.^{39–41}

The grading system was also shown to influence the learning response. This finding contrasts with other research,⁴² which showed that a pass/fail grading increased the time spent on studying, intrinsic motivation and achievement standards. Wilkinson⁴³ emphasised that the pass standard must be defined. Although criterion-referenced standards were used to designate performance as representative of 'fail', 'pass' or 'excellent' in the current setting, this was less clearly perceived by students in the context of the assessment task using pass/fail grading, which again indicates that careful implementation and the communication of expectations is important. As for motivational drive in a pass/fail setting, the perceptions of learners and responses to assessment and feedback information were shown to be very complex and dependent on the dynamics of personal and contextual features.²⁶ A student who is satisfied with a pass or expresses a need for a grade as an extrinsic motivator to learn is not necessarily a poor lifelong learner: other contextual factors may prevail at that time, such as the prioritising of another assessment or assignment tasks. There is some evidence that adding a (letter) grade to a portfolio-based system increases students' affirmation and self-assurance regarding knowledge and skills.⁴⁴ The combination of feedback and grades in a comprehensive programme of assessment needs to be carefully designed with awareness of the effects on student learning.

For a minority of students, the reflective practice in the portfolio as such was not perceived as useful; these students demonstrated limited intrinsic motivation and a deliberate effort to do only what was

necessary to pass the portfolio assessment. However, this was not always a black-and-white issue, and despite the extrinsic motivation, students indicated that they did learn from the information generated by the comprehensive programme of assessment. This aligns with the findings of Watling *et al.*,²⁶ who reported that the regulatory foci (promotion or prevention) to feedback may blend depending on the individual, task and context. To prevent such 'extrinsic' learners from entering a programme, the admission process could be improved by making sure that learning requirements are clearly communicated to candidates or perhaps even by adjusting non-cognitive selection criteria to include learning style. Further research is needed in this respect.

Limitations of the study

This study is subject to some limitations. Firstly, it was conducted in a single setting, the P-CI medical Master's degree course at Maastricht University, the Netherlands. Secondly, the first author (SH) had direct contact with the students in the context of the curriculum, which carried potential for a reflexive effect.¹⁹ Therefore, student interviews were carried out by the second author, who was not involved in the programme, and transcripts were anonymised. In addition, all steps were discussed with the research team and a logbook was kept. Thirdly, in this study, only students' introspections of their own constructions were studied. Given the interaction between teachers and students in assessment procedures, the perceptions of teachers are important to interpretation and to potential implications for practice.

In conclusion, the active participation of learners in their own learning was possible in and was supported by a comprehensive programme of assessment administered in combination with a portfolio. Students indicated that the methods employed and the procedures involved in the comprehensive programme of assessment did influence the construction of their learning, and that this influence could be either supportive or inhibiting. Important factors included the maintenance of balance in the frequency and timing of feedback and reflective activities, access to mentors, transparency in assessment procedures, attention to students who persist in using extrinsic motivation for reflective activities, clear communication of standards for grades, and the use of a purposeful combination of grades and feedback. Students did not acknowledge the formative features of assessment activities and hence this issue continues to represent a challenge in the setting of a comprehensive programme of

assessment. We found indications that this was partly linked to implementation issues, such as the use of marks which implicitly denoted a pass/fail judgement and the organisation of remediation sessions, which caused the end-of-block assessments to be seen as hurdles that needed to be passed and elicited pre-assessment effects on learning. Further research is needed to determine the conditions in which formative assessment or a low-stakes examination is truly interpreted as informative for learning. If this informative function is not feasible in practice, the theoretical model will have to be modified. More insight is required to resolve the fundamental question of whether or not the formative and summative functions of assessment can be combined. The merging of a comprehensive programme of assessment with reflective activities and coaching by a mentor resulted in a positive post-assessment learning effect whereby students self-regulate their (future) learning, despite continuing pre-assessment learning effects. Further research should now focus on how we can introduce more flexibility into assessment practices, and the conditions in which summative and formative aspects of assessment can best be combined and effectively used in a programmatic assessment approach.

Contributors: All authors contributed substantially to the conception and design of the study. SH and AOP acquired all data and took responsibility for the analysis and interpretation of the data in collaboration with LWTS, CPMvdV and EWD. SH wrote the manuscript in collaboration with LWTS, CPMvdV, AOP and EWD. All authors approved the final manuscript for submission and have agreed to be accountable for all aspects of the work.

Acknowledgement: None.

Funding: None.

Conflicts of interest: None.

Ethical approval: This study was approved by the Dutch Society for Medical Education (Nederlandse Vereniging voor Medische Onderwijs; no. NVMO-ERB-276).

REFERENCES

- Birenbaum M. Assessment preferences and their relationship to learning strategies and orientations. *High Educ* 1997;**33**:71–84.
- Cilliers FJ, Schuwirth L, Herman N, Adendorff HJ, van der Vleuten CPM. A model of the pre-assessment learning effects of summative assessment in medical education. *Adv Health Sci Educ Theory Pract* 2012;**17**:39–53.
- Dochy F, Segers M, Gijbels D, Struyven K. Assessment engineering. In: Boud D, Falchikov N, eds. *Rethinking Assessment in Higher Education: Learning for the Longer Term*. Oxford: Routledge 2007;87–100.
- Cilliers FJ, Schuwirth LWT, van der Vleuten CPM. Modelling the pre-assessment learning effects of assessment: evidence in the validity chain. *Med Educ* 2012;**46**:1087–98.
- Al-Kadri HM, Al-Moamary MS, Roberts C, van der Vleuten CPM. Exploring assessment factors contributing to students' study strategies: literature review. *Med Teach* 2012;**34** (Suppl):42–50.
- Gipps C. Assessment and learning. In: Gibbs C, ed. *Beyond Testing: Towards a Theory of Educational Assessment*. London: Farmer Press 2002;18–30.
- Shepard L. The role of assessment in a learning culture. *Educ Res* 2000;**29**:4–14.
- Schuwirth L, van der Vleuten C. Programmatic assessment: from assessment of learning to assessment for learning. *Med Teach* 2011;**33**:478–85.
- Bjork R, Dunlosky J, Kornell N. Self-regulated learning: beliefs, techniques, and illusions. *Annu Rev Psychol* 2013;**64**:417–44.
- Dijkstra J, van der Vleuten C, Schuwirth L. A new framework for designing programmes of assessment. *Adv Health Sci Educ Theory Pract* 2010;**15**:379–93.
- van der Vleuten C, Schuwirth L, Driessen E, Dijkstra J, Tigelaar D, Baartman L, van Tartwijk J. A model for programmatic assessment fit for purpose. *Med Teach* 2012;**34**:205–14.
- Dijksterhuis M, Schuwirth L, Braat D, Teunissen P, Scheele F. A qualitative study on trainees' and supervisors' perceptions of assessment for learning in postgraduate medical education. *Med Teach* 2013;**35**:e1396–402.
- Driessen E, Scheele F. What is wrong with assessment in postgraduate training? Lessons from clinical practice and educational research. *Med Teach* 2013;**35**:569–74.
- Dannefer EF, Henson L. The portfolio approach to competency-based assessment at the Cleveland Clinic Lerner College of Medicine. *Acad Med* 2007;**82**:493–502.
- Driessen E, van Tartwijk J, Govaerts M, Teunissen P, van der Vleuten C. The use of programmatic assessment in the clinical workplace: a Maastricht case report. *Med Teach* 2012;**34**:226–31.
- Bok H, Teunissen P, Favier RP, Rietbroek N, Theyse L, Brommer H, Haarhuis JC, van Beukelen P, van der Vleuten CP, Jaarsma DA. Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Med Educ* 2013;**13**:123.
- Bok HGJ, Teunissen PW, Spruijt A, Fokkema JPI, van Beukelen P, Jaarsma DADC, van der Vleuten CPM. Clarifying students' feedback-seeking behaviour in clinical clerkships. *Med Educ* 2013;**47**:282–91.
- Frank JR, ed. *The CanMEDS 2005 Physician Competency Framework. Better Standards. Better Physicians. Better Care*. Ottawa, ON: Royal College of Physicians and Surgeons of Canada 2005.
- Malterud K. Qualitative research: standards, challenges, and guidelines. *Lancet* 2001;**358**:483–8.

- 20 Bunniss S, Kelly DR. Research paradigms in medical education research. *Med Educ* 2010;**44**:358–66.
- 21 Onwuegbuzie AJ, Leech NL. A call for qualitative power analyses. *Qual Quant* 2007;**41**:105–21.
- 22 Pintrich P. A conceptual framework for assessing motivation and self-regulated learning in college students. *Educ Psychol Rev* 2004;**16**:385–407.
- 23 Archer JC. State of the science in health professional education: effective feedback. *Med Educ* 2010;**44**:101–8.
- 24 Hattie J, Timperley H. The power of feedback. *Rev Educ Res* 2007;**77**:81–112.
- 25 Shute VJ. Focus on formative feedback. *Rev Educ Res* 2008;**78**:153–89.
- 26 Watling C, Driessen E, van der Vleuten CPM, Vanstone M, Lingard L. Understanding responses to feedback: the potential and limitations of regulatory focus theory. *Med Educ* 2012;**46**:593–603.
- 27 Larsen DP, Butler AC, Roediger HL III. Test-enhanced learning in medical education. *Med Educ* 2008;**42**:959–66.
- 28 King N. Using templates in the thematic analysis of text. In: Cassel C, Symon G, eds. *Essential Guide to Qualitative Methods in Organizational Research*. London: Sage Publications 2004;256–70.
- 29 Bowen GA. Naturalistic inquiry and the saturation concept: a research note. *Qual Res* 2008;**8**:137–52.
- 30 Weston P, Smith C. The use of mini-CEX in UK foundation training six years following its introduction: lessons still to be learned and the benefit of formal teaching regarding its utility. *Med Teach* 2014;**36**:155–63.
- 31 Dolmans DHJM, de Grave W, Wolfhagen IHAP, Van der Vleuten CPM. Problem-based learning: future challenges for educational practice and research. *Med Educ* 2005;**39**:732–41.
- 32 Malik M, Diaz Voss Varela D, Stewart C, Laeeq K, Yenokyan G, Francis H, Bhatti NI. Barriers to implementing the ACGME Outcome Project: a systematic review of program director surveys. *J Grad Med Educ* 2012;**4**:425–33.
- 33 Fokkema JPI, Teunissen PW, Westerman M, van der Lee N, van der Vleuten CPM, Scherpbier AJJA, Dörr PJ, Scheele F. Exploration of perceived effects of innovations in postgraduate medical education. *Med Educ* 2013;**47**:271–81.
- 34 Al-Kadri HM, Al-Kadi MT, van der Vleuten CPM. Workplace-based assessment and students' approaches to learning: a qualitative inquiry. *Med Teach* 2013;**35** (Suppl):31–8.
- 35 Bennett R. Formative assessment: a critical review. *Assess Educ Princ Pol Pract* 2011;**18**:5–25.
- 36 Taras M. Summative assessment: the missing link for formative assessment. *J Furth High Educ* 2009;**33**:57–69.
- 37 Taras M. Assessment – summative and formative – some theoretical reflections. *Br J Educ Stud* 2005;**53**:466–78.
- 38 Norcini J, Anderson B, Bollela V et al. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. *Med Teach* 2011;**33**:206–14.
- 39 Tillema H. Portfolios as developmental assessment tools. *Int J Train Dev* 2001;**5**:126–35.
- 40 Buckley S, Coleman J, Davison I et al. The educational effects of portfolios on undergraduate student learning: a Best Evidence Medical Education (BEME) systematic review. BEME Guide No. 11. *Med Teach* 2009;**31**:282–98.
- 41 Mann K, Gordon J, MacLeod A. Reflection and reflective practice in health professions education: a systematic review. *Adv Health Sci Educ Theory Pract* 2009;**14**:595–621.
- 42 White C, Fantone J. Pass–fail grading: laying the foundation for self-regulated learning. *Adv Health Sci Educ Theory Pract* 2010;**15**:469–77.
- 43 Wilkinson T. Pass/fail grading: not everything that counts can be counted. *Med Educ* 2011;**45**:860–2.
- 44 Nowacki AS. Making the grade in a portfolio-based system: student performance and the student perspective. *Front Psychol* 2013;**4**:155.
- 45 Wrigley W, van der Vleuten C, Freeman A, Muijtjens A. A systemic framework for the progress test: strengths, constraints and issues: AMEE Guide No. 71. *Med Teach* 2012;**34**:68–97.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table S1. Interview question guide.

Table S2. Predefined codes used to guide analysis.

Received 15 June 2014 ; editorial comments to author 15 September 2014, accepted for publication 21 October 2014