

A new holistic way of assessment: programmatic assessment

Cees P.M. van der Vleuten, Sylvia Heeneman

Summary. Programmatic assessment is a holistic view on assessment since it considers the full assessment program as a focus. Deliberate choices are being made on methods of assessment, their arrangement in time and purpose. Each moment of assessment is considered to be one single data point. Any decision on a single data point is flawed and therefore pass/fail decisions are not being made. A single data point will only give quantitative or qualitative feedback relevant for learning. The follow-up of feedback, e.g. in learning objectives and reflective activities, learners are mentored by staff members. Decisions on learner progress are made when sufficient data points are being gathered and when the information is clear. Decision-making is done by committees and due process measures are taken to make these decisions robust. Programmatic assessment fits to a modern constructivist view on learning.

Key words. Assessment. Feedback. Learning. Programmatic assessment.

Una nueva forma holística de evaluación: la evaluación programática

Resumen. La evaluación programática es una visión integral de la evaluación, dado que considera como su objetivo el programa completo de evaluación. Se toman decisiones explícitas sobre los métodos de evaluación, su disposición en el tiempo y su propósito. Cada momento de la evaluación se considera que es un solo punto de información. Cualquier decisión basada en un solo dato de información es insuficiente y, por lo tanto, la decisión sobre si un alumno pasa o no pasa (aprueba o suspende) no se puede tomar únicamente en base a esto. Un único dato de información sólo proporcionará *feedback* cuantitativo o cualitativo relevante para el aprendizaje. En el proceso de seguimiento del *feedback*, por ejemplo respecto al cumplimiento de los objetivos de aprendizaje y de las actividades de reflexión, los docentes aconsejan a los estudiantes. Las decisiones sobre el progreso del alumno se hacen cuando se han recogido suficientes datos y cuando la información es clara. La toma de decisiones se realiza a través de comités y se toman las debidas medidas durante el proceso para que estas decisiones sean robustas. La evaluación programática se ajusta a la visión constructivista moderna en el aprendizaje.

Palabras clave. Aprendizaje. Evaluación. Evaluación programática. *Feedback*.

Introduction

To introduce programmatic assessment, a concrete example of an existing programme is given. In the graduate entry medicine program in Maastricht, students enter with a relevant bachelor degree. In 4 years' time they will do their medical studies and they will also receive a Master of Science degree in clinical research. The program itself is based on problem-based learning (PBL) in the first year, patient-based PBL [1] in the second year and 2 years of clinical rotations. In the fourth year the clinical rotation is one discipline of the preferred choice of the student and also contains a research period of nearly half a year. The admission is stringent; students are made clear they have to work hard and that we have high expectations of them. The intake is 50 students per year. The curriculum has been

structured according to the CanMEDS outcome framework [2], but there is a heavy emphasis on research and the scholarly role from that framework. The assessment is modular and block related, but also longitudinal. The block assessment is very diverse. It may contain classical MCT tests, open-ended questions, orals, project assignments, etcetera. A number of blocks have multiple dispersed mini-tests during a block. In the first 2 years there are a few OSCEs. During clinical work there is an elaborate program of work-based assessment (Mini-Cex, field notes, technical skills assessment, multi-source feedback) and we no longer use OSCEs here. The longitudinal assessment is cognitive and behavioural. The cognitive testing is through progress testing, a written test format that comprehensively assesses the end objectives of medical training [3]. It is administered to all students in the curriculum.

Department of Educational Development and Research (C.P.M. van der Vleuten). Department of Pathology (S. Heeneman). Maastricht University. Maastricht, The Netherlands.

Correspondencia: Prof. Cees van der Vleuten. Director School of Health Professions Education. Faculty of Health, Medicine and Life Sciences. Maastricht University. P.O. Box 616. 6200 MD Maastricht, The Netherlands.

E-mail: c.vandervleuten@maastrichtuniversity.nl

Recibido: 03.10.16.

Aceptado: 05.10.16.

Conflicto de intereses: No declarado.

Competing interests: None declared.

© 2016 FEM

For every new occasion a new test is developed. The behavioural assessment involves periodic peer and tutor assessment on the CanMEDS competencies in all years of the program. Every assessment provides information rich feedback, either quantitative (e.g. progress test) or qualitative (peer assessment, work-based assessments). Furthermore, no pass fail decisions are being taken in each individual assessment. All assessment information is gathered in an e-portfolio. Every student has a mentor who helps and guides the students throughout the whole 4 years of the program. Mentors are regular faculty members (clinicians or basic scientists) who are being trained for their role. Decision-making on student progress is based on the portfolio and is carried out by a committee at the end of the year. Mentors have no say in this decision. With a pass all ECTS credits of that year are awarded to the learner. During the year, a formative pass/fail is given through an evaluation by another (not the student's own) mentor.

A justification of its components

The illustration above shows what programmatic assessment entails in a nutshell. It contains of a number essential ingredients, which are going to be described systematically and for which a justification will be given.

Decoupling of assessment and pass/fail decisions

It is very easy to demonstrate how any single assessment is flawed. For example, regardless of the method, whether objective or subjective, reliability studies have indicated that we need some 3 to 4 hours of testing time to achieve minimal reliability and sometimes even more [4]. Most of our tests in actual practice have insufficient reliability and we make substantial false positive and negative decisions as a result of that. Similarly, any method has limitations in what it may measure and its validity will be limited. The practical implication is that one measure is really no measure and that we need to combine information as much as possible. In programmatic assessment any single assessment is called a data point. A data point is metaphorically similar to a pixel in a picture. A single pixel will not tell you what the picture is about. But many pixels will. So will many data points. This has been the reason to remove decisions from individual data points. The function of an individual data point is not to take a decision but to provide information meaningful to learning. This leads to the next critical element.

Individual data points are informative

Most assessment practices are relatively information poor. What typically is done is to give grades to learners. A grade is a very poor information carrier and kind of represents the poorest feedback one may get [5]. This is particularly true when complex skills are being assessed such as academic writing, or communication or professionalism. With these complex skills, grades are virtually meaningless and provide no cues for further improvement. Many of our assessment practices are therefore rather reductionist, very much focussed towards summative decision-making and ignoring feedback. In programmatic assessment every data point is information rich and feedback oriented. There should be no assessment without feedback. Feedback may be quantitative, for example profile scores may be given to learners on blueprint categories of a test being referenced to the performance of all learners. For our progress test we have developed an online feedback module in which the learner can see how his knowledge grows in time in any discipline, in any organ system category or in any aggregate of major clusters (basic, clinical behavioural sciences) and how that growth relates to all learners in the cohort [6]. Feedback may also be qualitative or narrative, for example comments from a clinical supervisor after a patient consultation indicating what went well, what needs improvement and what actions need to be taken. The recent literature becomes quite clear how narrative information provides a lot more information value than quantitative information does [7]. The assessment community is discovering this and we are shifting from scores to words [8]. The fact that decision making is not needed in individual data points allows the assessor not to worry about issues of subjectivity or reliability. The only concern is to provide rich information. A similar freedom is given to the choice of the method of assessment.

Eclectic choice of methods

In programmatic assessment constructive alignment is key to any data point: the method should reflect the intent of the instruction goals of the curriculum as closely as possible [9]. The choice of method will be defined by your justification for using it at that moment in time and in relationship to the programme of assessment as a whole [10]. Any method may go: traditional, authentic, subjective, individualized or team oriented. It is wise to vary methods of assessment to achieve desirable educa-

tional effects. You may have good reasons for having learners to write first, then to verbally express themselves, to synthesize information on the spot or to show daily real life clinical performance through a video or any other form of direct observation. All of these methods are good, provided the educational justification is good and the resulting information is meaningful to learning and drives the learning in a desirable way. To achieve maximal constructive alignment the educational task may also be the assessment task. For example, to be able to write an evidence-based medicine (EBM) synthesis of a clinical problem, a task scheduled in a learning program, but at the same time the quality of this task may be assessed. It is wise to design the programme of assessment as a whole so that deliberate choices can be made about complimentary methods or to ensure that a logical longitudinal build-up of skills being assessed. Classically our assessment methods are very modular. We typically end a course, block or module with some form of assessment. Think also of longitudinal assessment. Modern competency frameworks, such as the CanMEDS or any other framework, typically require curricula to have longitudinal strands throughout the program. The assessment may also be longitudinal in nature.

Feedback, reflection and self-directed learning needs support

The provision of feedback is not enough for feedback to be used [11]. Similarly, reflection as a basis for self-directed learning needs external support [12]. Therefore we have introduced a mentoring system where students are being coached through their training program. Mentoring has been shown to be a very powerful instrument for learner success and development [13]. The mentor has access to the e-portfolio. Mentor and learner meet a number of times throughout the year or in any other frequency that they deem important. Mentor meetings are prepared by learners. They are required to reflect on the information in the portfolio, to self-diagnose and to suggest potential remediation. Both learners and mentors appreciate their relationship. Learners are not anonymous persons in a big and challenging course and mentors cherish the close interaction with learners. Problems with learners, academic or personal, are spotted early on. Learners feel supported and they are challenged to excel. Minimum performance or disengagement is simply not tolerated. Mentors are being trained for their role, but more importantly they meet on a regular basis to exchange information and learn

themselves during these meetings, stressing the importance of a mentor-network.

Stakes of decision-making and number of data points are proportionally related

In programmatic assessment the notion of formative versus summative assessment is replaced by a continuum of stakes. Any individual data point is low stake. It is not of no stake, because any piece of information may be used in the whole process. Once there are sufficient pixels to understand the picture, high stake decisions can be taken. For example, promotion to the next year is a high stake decision. Many data points are needed to inform that decision. The many number of data points and the richness of information will form a solid basis for taking such a high stake decision. High stake decisions should be of no surprise to the learner. Therefore, intermediate decisions should be given as well. To make this affordable we have chosen that another mentor, foreign to the student, judges the portfolio half way during a year and gives a formative pass/fail.

Meaningful aggregation of assessment information

In arriving at a decision, all assessment information needs to be aggregated. Conventionally aggregation is done within a method to a total. For example, in an OSCE it is common aggregate information on a resuscitation station with a history taking and communication station. Yet these stations have conceptually little in common. In programmatic assessment information is aggregated across methods within meaningful categories. For example, the information on communication in the OSCE may be aggregated with information from a multisource feedback assessment. This also reveals the importance of structuring all assessment instruments in such a way that meaningful aggregation can be done. In practice this means that (most) assessments are structured according to competencies. When combining information simple 'averaging' is impossible. The information that is to be judged may be of a quantitative or qualitative nature. The aggregation needs a professional human judgment. This potentially leads to a subjective and flawed decision and this needs to be prevented.

High stake decision-making is procedurally robust

Instead of trying to make every data point objective, the collective of data points should be objective. This

is not achieved by for example strict use of checklists, but rather by taking procedural measures of due process that will bring credibility to the high-stake judgment [14]. A few examples may help. The decision is made by a committee of experts, not by an individual. The committee has independence to the mentors and the mentor has no say. This also protects the relationship between mentor and learner. The learner can be frank to the mentor. The committee uses narrative criteria to judge the portfolio. The criteria are narratives not checks. Checks would invite the process to trivialize. The narratives leave room for interpretation and provide flexibility. The committee deliberation depends on the clarity of the information. A lot of deliberation will take place when the information is unclear and when a pass/fail decision is critical. Critical decisions are justified by argumentation. The decision itself will hardly ever be a surprise to the learner due to the previous feedback cycles in the process. The learners may appeal to the decision if they feel that is appropriate. Appeals are dealt with by the examination committee. Periodically the decision-making process is being evaluated. Certain interesting cases may reveal shortcomings in the criteria. Or the experiences of decision-making may lead to an additional training of the assessors. All these measures will make decision robust and credible. Contrary to what might be expected, this assessment procedure is not very expensive. For 95% of learners the information is clear and the decision clear-cut. These meetings are well prepared and assessor time is only used when it is needed.

Discussion

Programmatic assessment differs dramatically with our traditional approach to assessment. Our traditional approach to assessment matches a traditional view on education: education is modular and mastery of every module is evidence of being competent. Showing mastery at the end of module test is sufficient and information may be quickly forgotten. The integration of knowledge or transfer to practice is left to the learner. This matches a mastery oriented learning approach or a behaviourist perspective on learning. Teacher-centred consumptive and inactive learning matches this education view. Modern education programmes are more constructivist. Knowledge and skills are constructed by learners. Learner-centred and active learning is the predominant approach. Much attention is given to transfer of knowledge to practice by intro-

ducing authentic learning task and early exposure to clinical practice. Complex skills are being addressed beyond the knowledge component. Learning is developmental, not compartmentalized. Programmatic assessment completely fits to this modern view on learning.

Programmatic assessment is also in line with all insights stemming from the evidence on assessment [15] and has been developed in a logical sequence of assessment insights [16]. It has been an attempt to reverse the traditional adagium 'assessment drives learning' towards 'learning drives assessment'. At the same time, it is clear that the radical different approach is difficult to implement. It requires quite some buy-in and understanding from the stakeholders involved. This is not an easy task. Such a change may be compared to changing a traditional curriculum to a problem-based curriculum. This also has been a tremendous hurdle for many organizations. Programmatic assessment may be used in undergraduate and postgraduate learning. The assessment concept is up taken by quite a few schools within medicine [17,18] and outside medicine in other health fields [19,20].

Naturally programmatic assessment needs thorough research on its value and effectiveness. The first research findings show positive results [19,21]. When progress tests results are compared between our graduate entry students and our regular undergraduate 6-year programme students, large performance differences are found in favour of the graduate entry students [22]. Learners become much more active feedback seekers and they really self-direct their learning. More research is needed and is ongoing. A few experiences so far are worth mentioning.

The implementation of programmatic assessment is a challenge as said before. It requires a major overhaul of the assessment program in which many stakeholders need to be convinced. Just like any other major educational change this requires an intensive change management strategy. Getting good feedback in the assessment process is a second challenge. Giving high quality feedback is a skill that needs to be developed. Faculty training is imperative. The mantra 'less is more' also holds here: less frequent high quality feedback is preferred over frequent low quality feedback. Actually poor quality feedback is less credible and incredible feedback is ignored by the learner [23]. Interestingly, the decision-making element in programmatic assessment is not so problematic. The procedures really works well in actual practice and appeals hardly occur. Cost might be another issue in programmatic assessment. Mentoring, individualized

feedback and committee-based decisions requires the necessary resources. The challenge is to carefully re-orientate resources. Our current assessment practices are expensive as well. Programmatic assessment requires a redistribution of assessment costs and this may require some sharp choices in what to discontinue in our current practice [24]. Our graduate entry program is state funded and there are no other sources for running it. It shows programmatic assessment is a viable assessment strategy. Finally, just like in problem-based learning hybrid implementations might be possible. Introducing programmatic assessment in the workplace seems somewhat easier than in school-based implementations. We have now redesigned our last 3 clinical years of the undergraduate curriculum (340 students per year) into a programmatic assessment form. Partial implementations may also be possible for example by introducing more feedback into an assessment program, or to introduce longitudinal monitoring of students or a mentoring system. Just like in problem-based learning [25], hybrid implementations will provide hybrid outcomes; full implementations will have the best chance of success.

Programmatic assessment optimizes both the learning function of assessment and the decision-making function. The richness of the pixels will be beneficial to the learning process and the collection of pixels will allow robust decision-making on learner progress. Programmatic assessment has the potential to harmonize assessment with modern constructivist approaches to learning.

References

1. Diemers AD, Dolmans DH, Van Santen M, Van Luijk SJ, Janssen-Noordman AM, Scherpbier AJ. Students' perceptions of early patient encounters in a PBL curriculum: a first evaluation of the Maastricht experience. *Med Teach* 2007; 29: 135-42.
2. Frank JR, Danoff D. The CanMEDS initiative: implementing an outcomes-based framework of physician competencies. *Med Teach* 2007; 29: 642-7.
3. Wrigley W, Van der Vleuten CP, Freeman A, Muijtjens A. A systemic framework for the progress test: strengths, constraints and issues: AMEE Guide No. 71. *Med Teach* 2012; 34: 683-97.
4. Van der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. *Med Educ* 2005; 39: 309-17.
5. Shute VJ. Focus on formative feedback. *Review of Educational Research* 2008; 78: 153-89.
6. Muijtjens AM, Timmermans I, Donkers J, Peperkamp R, Medema H, Cohen-Schotanus J, et al. Flexible electronic feedback using the virtues of progress testing. *Med Teach* 2010; 32: 491-5.
7. Ginsburg S, Eva K, Regehr G. Do in-training evaluation reports deserve their bad reputations? A study of the reliability and predictive ability of ITER scores and narrative comments. *Acad Med* 2013; 88: 1539-44.
8. Govaerts MJB, Van der Vleuten CPM. Validity in work-based assessment: expanding our horizons. *Med Educ* 2013; 47: 1164-74.
9. Biggs JB. Enhancing teaching through constructive alignment. *Higher Education* 1996; 32: 347-64.
10. Van der Vleuten CP, Schuwirth LW, Driessen EW, Dijkstra J, Tigelaar D, Baartman LK, et al. A model for programmatic assessment fit for purpose. *Med Teach* 2012; 34: 205-14.
11. Harrison CJ, Könings KD, Schuwirth L, Wass V, Van der Vleuten C. Barriers to the uptake and use of feedback in the context of summative assessment. *Adv Health Sci Educ* 2014; 48: 1-17.
12. Sargeant JM, Mann KV, Van der Vleuten CP, Metsemakers JF. Reflection: a link between receiving and using assessment feedback. *Adv Health Sci Educ* 2009; 14: 399-410.
13. Driessen EW, Overeem K. Mentoring. In Walsh K, ed. *Oxford textbook of medical education*. Oxford: Oxford University Press; 2013. p. 265-84.
14. Driessen E, Van der Vleuten C, Schuwirth L, Van Tartwijk J, Vermunt J. The use of qualitative research criteria for portfolio assessment as an alternative to reliability evaluation: a case study. *Med Educ* 2005; 39: 214-20.
15. Van der Vleuten CP, Schuwirth LW, Scheele F, Driessen EW, Hodges B. The assessment of professional competence: building blocks for theory development. *Best Pract Res Clin Obstet Gynaecol* 2010; 24: 703-19.
16. Vleuten CP. Revisiting 'Assessing professional competence: from methods to programmes'. *Med Educ* 2016; 50: 885-8.
17. Dannefer EF. Beyond assessment of learning toward assessment for learning: educating tomorrow's physicians. *Med Teach* 2013; 35: 560-3.
18. Chan T, Sherbino J; McMAP Collaborators. The McMaster Modular Assessment Program (McMAP): a theoretically grounded work-based assessment system for an emergency medicine residency program. *Acad Med* 2015; 90: 900-5.
19. Bok HG, Teunissen PW, Favier RP, Rietbroek NJ, Theyse LF, Brommer H, et al. Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Med Educ* 2013; 13: 123.
20. Palermo C, Gibson SJ, Dart J, Whelan K, Hay M. Programmatic assessment of competence in dietetics: a new frontier. *J Acad Nutr Diet* 2016; May 6. [Epub ahead of print].
21. Heeneman S, Oudkerk Pool A, Schuwirth LW, Vleuten CP, Driessen EW. The impact of programmatic assessment on student learning: theory versus practice. *Med Educ* 2015; 49: 487-98.
22. Heeneman S, Schut S, Donkers J, Van der Vleuten C, Muijtjens A. Embedding of the progress test in an assessment program designed according to the principles of programmatic assessment. *Med Teach* 2016; Sep 19 [Epub ahead of print].
23. Watling C, Driessen E, Van der Vleuten CP, Lingard L. Learning from clinical work: the roles of learning cues and credibility judgements. *Med Educ* 2012; 46: 192-200.
24. Van der Vleuten CPM, Heeneman S. On the issue of costs in programmatic assessment. *Perspect Med Educ* 2016; 5: 303-7.
25. Frambach JM, Driessen EW, Beh P, Van der Vleuten CP. Quiet or questioning? Students' discussion behaviors in student-centered education across cultures. *Studies in Higher Education* 2014; 39: 1001-21.