TWELVE TIPS

Twelve Tips for programmatic assessment

C.P.M. VAN DER VLEUTEN[1], L.W.T. SCHUWIRTH[2], E.W. DRIESSEN[1], M.J.B. GOVAERTS[1] &
S. HEENEMAN[1]

[1]Maastricht University, Maastricht, The Netherlands, [2]Flinders University, Adelaide, Australia

Abstract

Programmatic assessment is an integral approach to the design of an assessment program with the intent to optimise its learning function, its decision-making function and its curriculum quality-assurance function. Individual methods of assessment, purposefully chosen for their alignment with the curriculum outcomes and their information value for the learner, the teacher and the organisation, are seen as individual data points. The information value of these individual data points is maximised by giving feedback to the learner. There is a decoupling of assessment moment and decision moment. Intermediate and high-stakes decisions are based on multiple data points after a meaningful aggregation of information and supported by rigorous organisational procedures to ensure their dependability. Self-regulation of learning, through analysis of the assessment information and the attainment of the ensuing learning goals, is scaffolded by a mentoring system. Programmatic assessment-for-learning can be applied to any part of the training continuum, provided that the underlying learning conception is constructivist. This paper provides concrete recommendations for implementation of programmatic assessment.

Introduction

From the notion that every individual assessment has severe limitations in any criterion of assessment quality (van der Vleuten 1996), we proposed to optimise the assessment at the programme level (van der Vleuten & Schuwirth 2005). In a programme of assessment, individual assessments are purposefully chosen in such a way that the whole is more than the sum of its parts. Not every individual assessment, therefore, needs to be perfect. The dependability and credibility of the overall decision relies on the combination of the emanating information and the rigour of the supporting organisational processes. Old methods and modern methods may be used, all depending on their function in the programme as a whole. The combination of methods should be optimal. After the introduction of assessment programmes we have published conceptual papers on it (Schuwirth & van der Vleuten 2011, 2012) and a set of guidelines for the design of programmes of assessment (Dijkstra et al. 2012). More recently we proposed an integrated model for programmatic assessment that optimised both the learning function and the decision-making function in competency-based educational contexts (van der Vleuten et al. 2012), using well-researched principles of assessment (van der Vleuten et al. 2010). Whereas the Dijkstra et al. guidelines are generic in nature and even apply to assessment programmes without a curriculum (e.g. certification programmes), the integrated model is specific to constructivist learning programmes. In programmatic assessment decisions are decoupled from individual assessment moments. These individual assessment moments primarily serve for gathering information on the learner. Decisions are only made when sufficient information is gathered across individual moments of assessment. Programmatic assessment also includes a longitudinal view of learning and assessment in relation to certain learning outcomes. Growth and development is monitored and mentored. Decision-making on aggregated information is done by an (independent) group of examiners. Although this model of programmatic assessment is well received in educational practice (Driessen et al. 2012; Bok et al. 2013), many find programmatic assessment complex and theoretical. Therefore, in this paper we will describe concrete tips to implement programmatic assessment.

Tip 1

Develop a master plan for assessment

Just like a modern curriculum is based on a master plan, programmatic assessment has to be based on such a master plan as well. Essential here is the choice for an overarching structure usually in the form of a competency framework. This is important since in programmatic assessment pass/fail decisions are not taken at the level of each individual assessment moment, but only after a coherent interpretation can be made across many assessment moments. An individual assessment can be considered as a single data point. The traditional dichotomy between formative and summative assessment is redefined as a continuum of stakes, ranging from low- to high-stakes decisions. The stakes of the decision and the richness of the information emanating from the data points are related, ensuring proportionality of the decisions: high-stake decisions require many data points. In order to meaningfully aggregate information across these data points

Correspondence: C. P. M. van der Vleuten, Department of Educational Development and Research, Maastricht University, Maastricht, The Netherlands. E-mail: c.vandervleuten@maastrichtuniversity.nl

an overarching structure is needed, such as a competency framework. Information from various data points can be combined to inform the progress on domains or roles in the framework. For example, information on communication from an objective structured Clinical examination (OSCE) may be aggregated with information on communication from several mini-clinical evaluation exercise (Mini-CEX) and a multisource-feedback tool.

The master plan should therefore also provide a mapping of data points to the overarching structure and to the curriculum. The choices for each method and its content are purposefully chosen with a clear educational justification for using this particular assessment in this part of the curriculum in this moment in time. Many competency frameworks empha-sise complex skills (collaboration, professionalism, communi-cation, etc.) that are essentially behavioural, and therefore require longitudinal development. They are assessed through direct observation in real-life settings, under unstandardised conditions, in which professional, expert judgement is impera-tive. Depending on the curriculum and the phase of study, the master plan will thus contain a variety of assessment contents, a mixture of standardised and non-standardised methods and the inclusion of modular as well as longitudinal assessment elements. For any choice, the contribution to the master plan and through this alignment with the curriculum and the intended learning processes is crucial. The master plan for the curriculum and the assessment is ideally one single master plan.

The resulting subjectivity from non-standardised assess-ment using professional judgement is something that can be dealt with in programmatic assessment in two ways. First, by sampling many contexts and assessors, because many sub-jective judgements provide a stable generalisation from the aggregated data (van der Vleuten et al. 1991). Second, because subjectivity can be dealt through bias-reduction strategies showing due process in the way decisions are reached. We will revisit these latter strategies later in Tip 6. Subjectivity is *not* dealt with by removing professional judgement from the assessment process, for example, by over-structuring the assessment.

Tip 2

Develop examination regulations that promote feedback orientation

Individual data points are optimised for providing information and feedback to the learner about the quality of their learning and not for pass/fail decisions. Pass–fail decisions should not be made on the basis of individual data points – as is often the case in traditional regulations. Examination regulations trad-itionally connect credits to individual assessments; this should be prevented in programmatic assessment. Research has shown that feedback is ignored in assessment regimes with a summative orientation (Harrison et al. 2013). Because lining credits to individual assessments raises their stake, learners will primarily orientate themselves on passing the test instead of on feedback reception and follow-up (Bok et al. 2013). Credit points should be linked only to high stake decisions, based on

many data points. In all communication and most certainly in examination regulations the low-stake nature of individual assessments should be given full reign.

Tip 3

Adopt a robust system for collecting information

In programmatic assessment, information about the learner is essential and massive information is gathered over time. Being able to handle this information flexibly is vital. One way of collecting information is through the use of (electronic) portfolios. Here, portfolios have a dossier function allowing periodic analyses of the student's competence development and learning goals. The (e-)portfolio should therefore serve three functions: (1) provide a repository of formal and informal assessment feedback and other learning results (i.e. assess-ment feedback, activity reports, learning outcome products, and reflective reports), (2) facilitate the administrative and logistical aspects of the assessment process (i.e. direct online loading of assessment and feedback forms via multiple platforms, regulation of who has access to which information and by connecting information pieces to the overarching framework), and (3) enable a quick overview of aggregated information (such as overall feedback reports across sources of information). User friendliness is vital. The (e-)portfolio should be easily accessible to whatever stakeholder who has access to it. Many e-portfolios are commercially available, but care should be taken to ensure that the structure and functionalities of these portfolios are sufficiently aligned with the require-ments of the assessment programme.

Tip 4

Assure that every low-stakes assessment provides meaningful feedback for learning

Information richness is the cornerstone of programmatic assessment. Without rich assessment information program-matic assessment will fail. Mostly, conventional feedback from assessments, that is, grades and pass/fail decisions, are poor information carriers (Shute 2008). Meaningful feedback may have many forms. One is to give out the test material after test administration with information on the correct or incorrect responses. In standardised testing, score reports may be used that provide more detail on the performance (Harrison et al. 2013), for example, by giving online information on the blueprint categories of the assessment done, or on the skill domains (i.e. in an OSCE), or longitudinal overview for progress test results (Muijtjens et al. 2010). Sometimes verbal feedback in or after the assessment may be given (Hodder et al. 1989). In unstandardised assessment, quantitative infor-mation usually stems from the rating scales being used. This is useful, but it also has its limitations. Feedback for complex skills is enhanced by narrative information (Govaerts et al. 2007). Narrative information may also enrich standardised assessment. For example, in one implementation of program-matic assessment narrative feedback is given to learners on

weekly open-ended questions (Dannefer & Henson 2007). Given the fact that putting a metric on things that are difficult to quantify may actually trivialise what is being assessed. Metrics such as grades often lead to unwanted side effects like grade hunting and grade inflation. Grades may unintentionally "corrupt" the feedback process. Some argue we should replace scores with words (Govaerts & van der Vleuten 2013), particularly in unstandardised situations where complex skills are being assessed such as in clinical workplaces. This is not a plea against scores. Scoring and metrics are fine particularly for standardised testing. This is a plea for a mindful use of metrics and words when they are appropriate to use in order to provide meaningful feedback.

Obtaining effective feedback from teachers, supervisors or peers can be a tedious process, because it is time and resource intensive. Considering resource-saving procedures is interesting (e.g. peer feedback or automatic online feedback systems), but ultimately providing good quality feedback will cost time and effort. Two issues should be kept in mind when thinking about the resources. In programmatic assessment, assessment and learning are completely intertwined (assessment *as* learning), so the time for teaching and assessment becomes rather blurred. Second, more infrequent good feedback is better than frequent poor feedback. Feedback reception is highly dependent on the credibility of the feedback (Watling et al. 2012), so the "less-is-more" principle really applies to the process of feedback giving. High-quality feedback should be the prime purpose of any individual data point. If this fails within the implementation, programmatic assessment will fail.

Tip 5

Provide mentoring to learners

Feedback alone may not be sufficient for learners to be heeded well (Hattie & Timperley 2007). Research findings clearly indicate that feedback, reflection, and follow-up on feedback are essential for learning and expertise development (Ericsson 2004; Sargeant et al. 2009). Reflection for the mere sake of reflection is not well received by learners, but reflection as a basis for discussion is appreciated (Driessen et al. 2012). Feedback should ideally be part of a (reflective) dialogue, stimulating follow-up on feedback. Mentoring is an effective way to create such a dialogue and has been associated with good learning outcomes (Driessen & Overeem 2013).

In programmatic assessment mentoring is used to support the feedback process and the feedback use. In a dialogue with an entrusted person, performance may be monitored, reflections shared and validated, remediation activities planned, and follow-up may be negotiated and monitored. This is the role of a mentor. The mentor is a regular staff member, preferably having some knowledge over the curriculum. Mentor and learner meet each other periodically. It is important that the mentor is able to create a safe and entrusted relationship. For that purpose the mentor should be protected in having a judgemental role in the decision-making process (Dannefer & Henson 2007). The mentor's function is to get the best out of the learner. In conventional assessment programmes, adherence to minimum standards can suffice for promotion and graduation. In

programmatic assessment individual excellence is the goal and the mentor is the key person to promote such excellence.

Tip 6

Ensure trustworthy decision-making

High-stakes decisions must be based on many data points of rich information, that is, resting on broad sampling across contexts, methods and assessors. Since this information rich material will be of both quantitative and qualitative nature, aggregation of information requires professional judgement. Given the high-stakes nature, such professional judgement must be credible or trustworthy. Procedural measures should be put in place that bring evidence to this trustworthiness. These procedural measures may include (Driessen et al. 2013):

– An appointment of an assessment panel or committee responsible for decision-making (pass–fail–distinction or promotion decisions) having access to all the information, for example, embedded in the e-portfolio. Size and expertise of the committee will matter for its trustworthiness.

– Prevention of conflicts of interest and ensuring independence of panel members from the learning process of individual learners.

– The use of narrative standards or milestones.

– The training of committee members on the interpretation of standards, for example, by using exceptional or unusual cases from the past for training purposes.

– The organisation of deliberation proportional to the clarity of information. Most learners will require very little time; very few will need considerable deliberation. A chair should prepare efficient sessions.

– The provision of justification for decisions with high impact, by providing a paper trail on committee deliberations and actions, that is, document very carefully.

– The provision of mentor and learner input. The mentor knows the learner best. To eliminate bias in judgement and to protect the relationship with the learner, the mentor should not be responsible for final pass–fail decisions. Smart mentor input compromises can be arranged. For example, a mentor may sign for the authenticity of the e-portfolio. Another example is that the mentor may write a recommendation to the committee that may be annotated by the learner.

– Provision of appeals procedures.

This list is not exhaustive, and it is helpful to think of any measure that would stand up in court, such as factors that provide due process in procedures and expertise of the professional judgement. These usually lead to robust decisions that have credibility and can be trusted.

Tip 7

Organise intermediate decision-making assessments

High-stakes decisions at the end of the course, year, or programme should never be a surprise to the learner. Therefore, provision of intermediate assessments informing

643

the learner and prior feedback on potential future decisions is in fact another procedural measure adding to the credibility of the final decision. Intermediate assessments are based on fewer data points than final decisions. Their stakes are in between low-stake and high-stake decisions. Intermediate assessments are diagnostic (how is the learner doing?), therapeutic (what should be done to improve further?), and prognostic (what might happen to the learner; if the current development continues to the point of the high-stake decision?). Ideally, an assessment committee provides all intermediate evaluations, but having a full committee assessing all students may well be a too resource-intensive process. Less costly compromises are to be considered, such as using subcommittees or only the chair of the committee to produce these evaluations, or having the full committee only looking at complex student cases and the mentors evaluating all other cases.

Tip 8

Encourage and facilitate personalised remediation

Remediation is essentially different from resits or supplemental examinations. Remediation is based on the diagnostic information emanating from the on-going reflective processes (i.e. from mentor meetings, from intermediate evaluations, and from the learner self) and is always personalised. Therefore, the curriculum must provide sufficient flexibility for the learner to plan and complete remediation. There is no need for developing (costly) remediation packages. Engage the learner in making decisions on what and how remediation should be carried out, supported by an experienced mentor. Ideally, remediation is made a responsibility of the learner who is provided with sufficient support and input to achieve this.

Tip 9

Monitor and evaluate the learning effect of the programme and adapt

Just like a curriculum needs evaluation in a plan-do-act-cycle, so does an assessment programme. Assessment effects can be unexpected, side effects often occur, assessment activities, particularly very routine ones, often tend to trivialise and become irrelevant. Monitor, evaluate, and adapt the assessment programme systematically. All relevant stakeholders involved in the process of programmatic assessment provide a good source of information on the quality of the assessment programme. One very important stakeholder is the mentor. Through the mentor's interaction with the learners, they will have an excellent view on the curriculum in action. This information could be systematically gathered and exchanged with other stakeholders responsible for the management of the curriculum and the assessment programme. Most schools will have a system for data-gathering on the quality of the educational programme. Mixed-method approaches combining quantitative and qualitative information are advised (Ruhe & Boudreau 2013). Similarly, learners should be able

to experience the impact of the evaluations on actual changes in the programme (Frye & Hemmer 2012).

Tip 10

Use the assessment process information for curriculum evaluation

Assessment may serve three functions: to promote learning, to promote good decisions on whether learning outcomes are achieved, and to evaluate the curriculum. In programmatic assessment, the information richness is a perfect basis also for curriculum evaluation. The assessment data gathered, for example, in the e-portfolio, not only provides an X-ray of the competence development of the learners, but also on the quality of the learning environment.

Tip 11

Promote continuous interaction between the stakeholders

As should be clear from the previous, programmatic assessment impacts at all levels: students, examiners, mentors, examination committees, assessment developers, and curriculum designers. Programmatic assessment is, therefore, the responsibility of the whole educational organisation. When implemented, frequent and on-going communication between the different stakeholder groups is essential in the process. Communication may regard imperfections in the operationalisation of standards or milestones, incidents, and interesting cases that could have consequences for improvement of the system. Such communication could eventually affect procedures and regulations and may support the calibration of future decisions. For example, a firewall between the assessment committee and mentors fosters objectivity and independency of the decision-making, but at the same time may also hamper information richness. Sometimes, however, decisions need more information about the learner and then continuous communication processes are indispensable. The information richness in programmatic assessment enables us to make the system as fair as possible.

Tip 12

Develop a strategy for implementation

Programmatic assessment requires a culture change in thinking about assessment that is not easy to achieve in an existing educational practice. Traditional assessment is typically modular, with summative decisions and grades at the end of modules. When passed, the module is completed. When failed, repetition through resits or through repetition of the module is usually the remedy. This is all very appropriate in a mastery learning view on learning. However, modern education builds on constructivist learning theories, starting from notions that learners create their own knowledge and skills, in horizontally and/or vertically integrated programmes to guide and support competence. Programmatic assessment is better aligned to notions of constructivist learning and longitudinal

competence development through its emphasis on feedback, use of feedback to optimise individual learning and remediation tailored to the needs of the individual student. This radical change often leads to fear that such assessment systems will be soft and vulnerable to gaming of students, whereas the implementation examples demonstrate the opposite effect (Bok et al. 2013). Nevertheless, for this culture change in assessment a change strategy is required, since many factors in higher education are resistant to change (Stephens & Graham 2010). A change strategy needs to be made at the macro-, meso- and micro levels.

At the macro level, national legal regulations and university regulations are often strict about assessment policies. Some universities prescribe grade systems to be standardised across all training programmes. These macro level limitations are not easy to influence, but it is important to know the "wriggle room" these policies leave for the desired change in a particular setting. Policy-makers and administrators need to become aware of why a different view on assessment is needed. They also need to be convinced on the robustness of the decision-making in an assessment programme. The qualitative ontology underlying the decision-making process in programmatic assessment is a challenging one in a positivist medical environment. Very important is to explain programmatic assessment in a language that is not jargonistic and which aligns with the stakeholder's professional language. For clinicians, for example, analogies with diagnostic procedures in clinical health care often prove helpful.

At the meso level programmatic assessment may have consequences for the curriculum. Not only should the assessment be aligned with the overarching competency framework, but with the curriculum as well. Essential are the longitudinal lines in the curriculum requiring a careful balance of modular and longitudinal elements. Individual stakeholders and committees need to be involved as early as possible. Examination rules and regulations need to be constructed which are optimally transparent, defensible, but which respect the aggregated decision-making in programmatic assessment. The curriculum also needs to allow sufficient flexibility for remediation. Leaders of the innovation need to be appointed, who have credibility and authority.

Finally, at the micro level teachers and learners need to be involved in the change right from the start. Buy-in from teachers and learners is essential. To create buy-in the people involved should understand the nature of the change, but more importantly they should be allowed to see how the change also addresses their own concerns with the current system. Typically, teaching staff do have the feeling that something in the current assessment system is not right, or at least suboptimal, but they do not automatically make the connection with programmatic assessment as a way to solve these problems.

The development of programmatic assessment is a learning exercise for all and it is helpful to be frank about unexpected problems to arise during the first phases of the implementation; that is innate to innovation. So it is therefore good to structure this learning exercise as a collective effort, which may exceed traditional faculty development (De Rijdt et al. 2013). Although conventional faculty development is needed,

involving staff and students in the whole design process supports the chance of success and the creation of ownership (Könings et al. 2005) and creates a community of practice promoting sustainable change (Steinert 2014).

Changing towards programmatic assessment can be compared with changing traditional programmes to problem-based learning (PBL). Many PBL implementations have failed due to problems in the implementation (Dolmans et al. 2005). When changing to programmatic assessment, careful attention should be paid to implementation and the management of change at all strategic levels.

Conclusion

Programmatic assessment has a clear logic and is based on many assessment insights that have been shaped trough research and educational practice. Logic and feasibility, however, are inversely related in programmatic assessment. To introduce full-blown programmatic assessment in actual practice all stakeholders need to be convinced. This is not an easy task. Just like in PBL, partial implementations are possible with programmatic assessment (i.e. the increase in feedback and information in an assessment programme, mentoring). Just like in PBL, this will lead to partial success. We hope these tips will allow you to get as far as you can get.

Notes on contributors

CEES VAN DER VLEUTEN, PhD, is Professor of Education, Scientific Director of the School of Health Professions Education, Faculty of Health, Medicine and Life Sciences, Maastricht University, the Netherlands.

LAMBERT SCHUWIRTH, MD, PhD, is Professor of Medical Education, Health Professions Education, Flinders Medical School, Adelaide, South Australia.

ERIK DRIESSEN, PhD, is Associate Professor, Chair of the Department of Educational Development and Research, Department of Educational Development and Research, Faculty of Health, Medicine and Life Sciences, Maastricht University, the Netherlands.

MARJAN GOVAERTS, MD, PhD, is Assistant Professor, Department of Educational Development and Research, Faculty of Health, Medicine and Life Sciences, Maastricht University, the Netherlands.

SYLVIA HEENEMAN, PhD, is Professor of Medical Education, Department of Pathology, Faculty of Health, Medicine and Life Sciences, Maastricht University, the Netherlands.

References

Bok HG, Teunissen PW, Favier RP, Rietbroek NJ, Theyse LF, Brommer H, Haarhuis JC, van Beukelen P, van der Vleuten CP, Jaarsma DA. 2013. Programmatic assessment of competency-based workplace learning: When theory meets practice. BMC Med Educ 13:123.

Dannefer EF, Henson LC. 2007. The portfolio approach to competency-based assessment at the Cleveland Clinic Lerner College of Medicine. Acad Med 82:493–502.

De Rijdt C, Stes A, van der Vleuten C, Dochy F. 2013. Influencing variables and moderators of transfer of learning to the workplace within the area

of staff development in higher education: Research review. Educ Res Rev 8:48–74.

Dijkstra J, Galbraith R, Hodges BD, McAvoy PA, McCrorie P, Southgate LJ, van der Vleuten CP, Wass V, Schuwirth LW. 2012. Expert validation of fit-for-purpose guidelines for designing programmes of assessment. BMC Med Educ 12:20.

Dolmans DH, De Grave W, Wolfhagen IH, van der Vleuten CP. 2005. Problem-based learning: Future challenges for educational practice and research. Med Educ 39:732–741.

Driessen EW, Heeneman S, van der Vleuten CPM. 2013. Portfolio assessment. In: Dent JA, Harden RMM, editors. A practical guide for medical teachers. Edinburgh: Elsevier Health Sciences. pp 314–323.

Driessen EW, Overeem K. 2013. Mentoring. In: Walsh K, editor. Oxford textbook of medical education. Oxford: Oxford University Press. pp 265–284.

Driessen EW, van Tartwijk J, Govaerts M, Teunissen P, van der Vleuten CP. 2012. The use of programmatic assessment in the clinical workplace: A Maastricht case report. Med Teach 34:226–231.

Ericsson KA. 2004. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. Acad Med 79: S70–S81.

Frye AW, Hemmer PA. 2012. Program evaluation models and related theories: AMEE Guide No. 67. Med Teach 34:e288–e299.

Govaerts MJB, van der Vleuten CPM. 2013. Validity in work-based assessment: Expanding our horizons. Med Educ 47:1164–1174.

Govaerts MJ, van der Vleuten CP, Schuwirth LW, Muijtjens AM. 2007. Broadening perspectives on clinical performance assessment: Rethinking the nature of in-training assessment. Adv Health Sci Educ 12:239–260.

Harrison CJ, Könings KD, Molyneux A, Schuwirth LW, Wass V, van der Vleuten CP. 2013. Web-based feedback after summative assessment: How do students engage? Med Educ 47:734–744.

Hattie J, Timperley H. 2007. The power of feedback. Rev Educ Res 77: 81–112.

Hodder RV, Rivington R, Calcutt L, Hart I. 1989. The effectiveness of immediate feedback during the objective structured clinical examination. Med Educ 23:184–188.

Könings KD, Brand-Gruwel S, Merriënboer JJ. 2005. Towards more powerful learning environments through combining the perspectives of designers, teachers, and students. Br J Educ Psychol 75:645–660.

Muijtjens AM, Timmermans I, Donkers J, Peperkamp R, Medema H, Cohen-Schotanus J, Thoben A, Wenink AC, van der Vleuten CP. 2010. Flexible electronic feedback using the virtues of progress testing. Med Teach 32: 491–495.

Ruhe V, Boudreau JD. 2013. The 2011 program evaluation standards: A framework for quality in medical education programme evaluations. J Eval Clin Pract 19:925–932.

Sargeant JM, Mann KV, van der Vleuten CP, Metsemakers JF. 2009. Reflection: A link between receiving and using assessment feedback. Adv Health Sci Educ: Theory Pract 14:399–410.

Schuwirth LWT, van der Vleuten CPM. 2011. Programmatic assessment: From assessment of learning to assessment for learning. Med Teach 33: 478–485.

Schuwirth LW, van der Vleuten CP. 2012. Programmatic assessment and Kane's validity perspective. Med Educ 46:38–48.

Shute VJ. 2008. Focus on formative feedback. Rev Educ Res 78: 153–189.

Steinert Y. 2014. Medical education and faculty development. The Wiley Blackwell Encyclopedia of Health, Illness, Behavior, and Society. pp 1344–1348.

Stephens JC, Graham AC. 2010. Toward an empirical research agenda for sustainability in higher education: Exploring the transition management framework. J Cleaner Prod 18:611–618.

van der Vleuten CPM. 1996. The assessment of professional competence: Developments, research and practical implications. Adv Health Sci Educ 1:41–67.

van der Vleuten CP, Schuwirth LW. 2005. Assessing professional competence: From methods to programmes. Med Educ 39:309–317.

van der Vleuten CP, Norman GR, De Graaff E. 1991. Pitfalls in the pursuit of objectivity: Issues of reliability. Med Educ 25:110–118.

van der Vleuten CP, Schuwirth LW, Driessen EW, Dijkstra J, Tigelaar D, Baartman LK, van Tartwijk J. 2012. A model for programmatic assessment fit for purpose. Med Teach 34:205–214.

van der Vleuten CP, Schuwirth LW, Scheele F, Driessen EW, Hodges B. 2010. The assessment of professional competence: Building blocks for theory development. Best Pract Res Clin Obstet Gynaecol 24: 703–719.

Watling C, Driessen E, van der Vleuten CP, Lingard L. 2012. Learning from clinical work: The roles of learning cues and credibility judgements. Med Educ 46:192–200.