

Ottawa 2020 consensus statement for programmatic assessment – 1. Agreement on the principles

Sylvia Heeneman, Lubberta H. de Jong, Luke J. Dawson, Tim J. Wilkinson, Anna Ryan, Glendon R. Tait, Neil Rice, Dario Torre, Adrian Freeman & Cees P. M. van der Vleuten

To cite this article: Sylvia Heeneman, Lubberta H. de Jong, Luke J. Dawson, Tim J. Wilkinson, Anna Ryan, Glendon R. Tait, Neil Rice, Dario Torre, Adrian Freeman & Cees P. M. van der Vleuten (2021): Ottawa 2020 consensus statement for programmatic assessment – 1. Agreement on the principles, Medical Teacher, DOI: [10.1080/0142159X.2021.1957088](https://doi.org/10.1080/0142159X.2021.1957088)

To link to this article: <https://doi.org/10.1080/0142159X.2021.1957088>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 03 Aug 2021.



Submit your article to this journal [↗](#)



Article views: 1237



View related articles [↗](#)







View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)

Ottawa 2020 consensus statement for programmatic assessment – 1. Agreement on the principles

Sylvia Heeneman^a, Lubberta H. de Jong^b, Luke J. Dawson^c, Tim J. Wilkinson^d , Anna Ryan^e,
Glendon R. Tait^f, Neil Rice^g , Dario Torre^h , Adrian Freeman^g and Cees P. M. van der Vleutenⁱ 

^aDepartment of Pathology, School of Health Profession Education, Maastricht University, Maastricht, The Netherlands; ^bDepartment of Population Health Sciences, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands; ^cSchool of Dentistry, University of Liverpool, Liverpool, UK; ^dEducation Unit, University of Otago, Christchurch, New Zealand; ^eDepartment of Medical Education, Melbourne Medical School, University of Melbourne, Melbourne, Australia; ^fMD Program, Department of Psychiatry, and The Wilson Centre, University of Toronto, Toronto, Canada; ^gCollege of Medicine and Health, University of Exeter Medical School, Exeter, UK; ^hDepartment of Medicine, Uniformed Services University of Health Sciences, Bethesda, MD, USA; ⁱDepartment of Educational Development and Research, School of Health Profession Education, Maastricht University, Maastricht, The Netherlands

ABSTRACT

Introduction: In the Ottawa 2018 Consensus framework for good assessment, a set of criteria was presented for systems of assessment. Currently, programmatic assessment is being established in an increasing number of programmes. In this Ottawa 2020 consensus statement for programmatic assessment insights from practice and research are used to define the principles of programmatic assessment.

Methods: For fifteen programmes in health professions education affiliated with members of an expert group ($n=20$), an inventory was completed for the perceived components, rationale, and importance of a programmatic assessment design. Input from attendees of a programmatic assessment workshop and symposium at the 2020 Ottawa conference was included. The outcome is discussed in concurrence with current theory and research.

Results and discussion: Twelve principles are presented that are considered as important and recognisable facets of programmatic assessment. Overall these principles were used in the curriculum and assessment design, albeit with a range of approaches and rigor, suggesting that programmatic assessment is an achievable education and assessment model, embedded both in practice and research. Knowledge on and sharing how programmatic assessment is being operationalized may help support educators charting their own implementation journey of programmatic assessment in their respective programmes.

KEYWORDS

Programmatic assessment;
curriculum; feedback;
decision-making;
teachers; learners

Background

In 2010, the Ottawa conference produced a set of consensus criteria for good assessment (Norcini et al. 2011). It was recognised that a similar set of criteria would be needed for systems of assessment, which goes beyond single assessments, and systematically combines a series of individual measures that are subsequently integrated to provide evidence for a certain purpose, e.g. a decision for graduation or promotion to a subsequent year. Therefore, in the Ottawa 2018 consensus framework, a separate framework applying to systems of assessment was presented (Norcini et al. 2018). As described in the Ottawa 2018 consensus, systems of assessment can have various formats. A system can consist of a series of assessments, combined with other information, to facilitate a multi-layered decision, e.g. admission and licensure systems. Other systems of assessment prioritise educational and instructional design approaches, such as progress testing and programmatic assessment (Norcini et al. 2018).

Programmatic assessment was introduced by van der Vleuten and Schuwirth (van der Vleuten et al. 2012; van

der Vleuten and Schuwirth 2005) and is based on the principle that every individual assessment method or tool has limitations and compromises are needed if just individual assessments are used for (pass–fail) decisions. In contrast, common assessment approaches are often modular, with an end of period/module/course assessment, that leads to a grade and an associated pass–fail decision. This traditional summative approach to assessment has multiple unintended consequences, such as driving undesirable learning approaches, promoting extrinsic motivation, and ignoring any feedback that is given (van der Vleuten and Schuwirth 2005). The programmatic assessment model as proposed by van der Vleuten and Schuwirth, is a potential solution to the abovementioned problems. The programmatic assessment model has been defined as a specific approach to the design of assessment and education aimed at optimising the learning and decision function of assessment. Assessment information and feedback, originating from multiple data points in a variety of assessment formats, is aggregated by the learner and staff and is used for learning and for high-stakes decisions such as

CONTACT Sylvia Heeneman  s.heeneman@maastrichtuniversity.nl  Department of Pathology, School of Health Profession Education, Maastricht University, Peter Debyealaan 52, 6229 ER Maastricht, The Netherlands

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

promotion to the next year or certification (Schuwirth and van der Vleuten 2011; van der Vleuten et al. 2015).

Programmatic assessment is built on a number of key principles, as outlined in various key papers (Table 1). It is however important to realise that programmatic assessment is an instructional design approach (van der Vleuten and Schuwirth 2005) and its acceptability is strongly influenced by a variety of factors such as the values of the educational programme and limitations imposed by institutional requirements. It is critical to note that programmatic assessment is an assessment concept and not a recipe. In the context of a conventional teacher-centred curriculum with a set of modules or courses that need be passed, a programmatic assessment approach has less value. In a learner-centred curriculum with a constructivist view on education, using longitudinal skill development and with an emphasis on life-long learning and self-directed learning, programmatic assessment is a natural fit. The principles as delineated in current literature (Table 1) are important, yet can be realised in many different manifestations. The way coaching is organised or the way in which high-stakes decisions are made, can vary and different choices can be made. In addition, programmes should design their educational and assessment practices fit-for-purpose given local contextual factors, and ensure alignment of education and assessment within the curriculum.

The key principles as delineated in current literature (Table 1) are important for defining whether the assessment and education approach should be characterised as programmatic assessment rather than 'programmes of assessment'. All schools have a programme of assessment, but not all are programmatic. To be programmatic, the theoretical principles should be integrated into the design of the teaching and the assessment programme, and principles pertaining to both the learning function (i.e. principle 1/2/3/11/12) and the decision function of assessment (i.e. principle 7/8/9/10) should be present. Currently, programmatic assessment is implemented in an increasing number of programmes and research evidence is accumulation around the effects of implementations on teachers, learners, and policies (Schut et al. 2021). The aim of this Ottawa 2020 consensus paper is to use insights from practice and research to define agreement on the principles for programmatic assessment, which are presented in Table 2.

The Ottawa 2020 consensus working group drafted two papers (part 1 and 2). In this paper (part 1), a consensus on the principles of programmatic assessment was sought via an inventory of the perceived components, rationale, and importance, collected from an expert group affiliated with educational programmes that use or have implemented various aspects of programmatic assessment, and through input from attendees in the programmatic assessment workshop and/or symposium at the Ottawa 2020 conference. In a second paper (part 2), after reaching consensus on the principles, additional data were collected to provide a practical implementation of the consensus principles and well as an understanding of how programmatic assessment was implemented across different educational and institutional contexts (Torre et al. 2021). The aim of both papers is to use insights from practice and research to reflect on the principles and practical realisations. Together these two papers may inform choices and contribute to the decision making of

health professional programmes that might consider implementing programmatic assessment.

Consensus on the theoretical principles of programmatic assessment

The theory and model of programmatic assessment as proposed by van der Vleuten and Schuwirth was the starting point for part 1 to reach a consensus on the principles of programmatic assessment. The first step was to assemble a number of experts that have experience with either the practice, such as programme leaders or directors of assessment that introduced or are using programmatic assessment in their programmes, or experience with theory and research, such as scholars and educationalists, or both. A convenience sampling method was used based on existing collaborations and conference networks of two of the authors (AF, CvdV). This led to a group of 20 experts affiliated to either undergraduate or post-graduate programmes, with the majority being in medical education. The group of affiliated experts decided a priori to use a set of principles as stated in the foundation papers describing the programmatic assessment model (Schuwirth and van der Vleuten 2011; van der Vleuten et al. 2012; van der Vleuten and Schuwirth 2005).

The second step was to make an inventory for the perceived components, rationale, and importance of the programmatic assessment design. The group of experts were invited to complete a survey with questions regarding the principles as shown in Table 1:

- i. *Do you agree with this principle?*
- ii. *What do you think are components within this principle?*
- iii. *How important is this principle in programmatic assessment?*
- iv. *Is this principle easy to adhere to?*
- v. *Have you implemented this principle in your programme and if yes, how?*

In a pilot prior to sending the invitation with the survey to the expert group, a subgroup (SH, LdJ, LD, TW) first completed the survey and after discussion decided to combine the responses for principles 1–3, principles 4 and 5, and principles 6 and 7, given that these share similar theoretical tenets of the programmatic assessment model. The survey was completed by experts representing 15 programmes from six countries across three continents. The characteristics of these 15 programmes are shown in Table 3.

A subgroup of the expert group (SH, TW, AR, LdJ, LdD, DT) did a first analysis to provide an overview of the data by categorising the responses for question i, iii, iv, and v into a scale of yes, partially, or no, taking into account the range of the individual response. In addition, a simple thematic analysis was undertaken on the narratives of all questions. This initial overview of the data was then discussed within the subgroup, to prepare for the pre-conference discussion with the expert group members at the Ottawa 2020 conference.

The analysis and discussion of the draft consensus with the members of the expert group present at the pre-

Table 1. Principles of programmatic assessment as based on literature, used as the starting point of the consensus process (Van der Vleuten et al. 2010, 2012, 2017, 2019; van der Vleuten and Schuwirth 2005).

1	Every (part of an) assessment is but a data-point
2	Every data-point is optimised for learning by giving meaningful feedback to the learner
3	Pass/fail decisions are not given on a single data-point
4	There is a mix of methods of assessment
5	The choice of method depends on the educational justification for using that method
6	The distinction between summative and formative is replaced by a continuum of stakes
7	Stakes and decision-making learner progress are proportionally related to the stakes
8	Assessment information is triangulated across data-points towards a competency framework
9	High-stakes decisions (promotion, graduation) are made in competence committees
10	Intermediate decisions are made with the purpose of informing the learner on their progress
11	Learners have recurrent learning meetings with (faculty) mentors using a self-analysis of all assessment data
12	The assessment is optimally bespoke or tailored to the individual learner

Table 2. Final Ottawa 2020 consensus principles of programmatic assessment after input of the expert group and Ottawa attendees, changes as compared to Table 1 are indicated in bold.

1	Every (part of an) assessment is but a data-point
2	Every data-point is optimised for learning by giving meaningful feedback to the learner
3	Pass/fail decisions are not given on a single data-point
4	There is a mix of methods of assessment
5	The method chosen should depend on the educational justification for using that method
6	The distinction between summative and formative is replaced by a continuum of stakes
7	Decision-making on learner progress is proportionally related to the stake
8	Assessment information is triangulated across data-points towards an appropriate framework
9	High-stakes decisions (promotion, graduation) are made by in a credible and transparent manner, using a holistic approach
10	Intermediate review is made to discuss and decide with the learner on their progression
11	Learners have recurrent learning meetings with (faculty) mentors/coaches using a self-analysis of all assessment data
12	Programmatic assessment seeks to gradually increase the learner's agency and accountability for their own learning through the learning being tailored to support individual learning priorities

Table 3. Characteristics of the programmes of members of the expert group, who completed the survey on agreement, importance, adherence, and implementation of the principles of programmatic assessment.

	Discipline	Undergraduate (UG) post-graduate/speciality training (PG)	Number of learners	Country	Reference, in which the programme is described, if available
1	Medicine	UG	280/year	United Kingdom	Freeman and Ricketts (2010)
2	Medicine	UG	32/year	USA	Dannefer and Henson (2007) and Dannefer and Prayson (2013)
3	Dentistry	UG	377/year	United Kingdom	Dawson et al. (2015)
4	Medicine	PG	38	Canada	Chan and Sherbino (2015)
5	Medicine	UG	300/year	New Zealand	Wilkinson et al. (2011)
6	Medicine	UG	50/year	Netherlands	Heeneman et al. (2015) and Schut et al. (2018)
7	Veterinary	UG	225/year	Netherlands	Bok et al. (2013) and de Jong et al. (2019)
8	Medicine	UG	260/year	Canada	Kulasegaram et al. (2018)
9	Nutrition	PG	20	Australia	Jamieson et al. (2017)
10	Medicine	UG	350/year	Netherlands	Partly in Driessen et al. (2012)
11	Medicine	UG	360/year	Australia	Ryan et al. (2017)
12	Medicine	UG	200/year	Australia	–
13	Medicine	PG	462	Australia	Schuwirth et al. (2017)
14	Medicine	UG	288/year	Canada	–
15	Medicine	PG	136	Canada	Laughlin et al. (2012)

conference discussion showed that language and formulation were important. The impact of language and rhetoric has been shown in other educational practices, such as the teaching of patient communication or the interpretation of the word 'competence' (Lingard 2007, 2009). The phrasing of principles 9 and 12 led to certain misunderstandings and needed further clarification and refinement. Consequently, principles 9 and 12 were rephrased and experts' responses to the five questions were then recollected and reanalysed, prior to the workshop at the conference. Thirteen of the 15 experts responded to this request.

The data for the categorisation of agreement, importance, adherence, and implementation are summarised in Table 4. The first analysis was presented at the Ottawa 2020 conference workshop, and the audience was asked to provide feedback and points for discussion. This feedback

process led to a further redefinition of principles 5, 7, 8, 10, and 11. For principle 8, it was discussed that the framework used for triangulation and aggregation is not necessarily a competency-based framework, therefore this was changed to an 'appropriate' framework (*principle 8: Assessment information is triangulated across data-points towards an appropriate framework*). For principle 10, the central role of the learner in the review of his/her performance data and the purpose of the intermediate review was discussed, and led to a change in the phrasing (*principle 10: Intermediate review is made with to discuss and decide with the learner on their progression*). For principle 11, the word mentor was supplemented by the word coach. For principles 5 and 7, the changes mainly concerned grammar and syntax. Together with the thematic analysis of the narratives and the feedback during the conference, this led to the following agreement on the principles:

Table 4. Overview of the data by categorising the responses for questions on agreement, importance, adherence, and implementation of the principles of programmatic assessment.

Principle	Agreement Yes/partially/no	Importance Yes/mixed/no	Easy to adhere Yes/partially/no	Implemented Yes/partially/no
1/2/3	14/1/0	10/4/1	8/0/7	13/2/0
4/5	15/0/0	9/6/0	12/0/2	12/2/1
6/7	15/0/0	6/5/4	7/0/8	13/0/2
8	14/1/0	8/4/3	9/0/5	12/2/1
9	15/0/0	7/6/2	11/3/0	13/2/0
9 ^a	13/0/0	12/0/0	12/1/0	13/0/0
10	13/1/1	8/5/2	9/0/4	13/1/1
11	12/3/0	9/3/3	10/0/5	10/4/1
12	7/5/1	2/4/8	5/0/9	8/3/3
12 ^a	11/2/0	10/2/0	5/6/1	8/3/0

^aThe responses to rephrased principle 9 and 12 were recollected after the first analysis and discussion with the experts, prior to the conference.

White and bold: majority of programs confirmed. Light grey and italics: programs were mixed in their responses. Dark grey and underlined: majority of programs did not confirm.

Principle 1/2/3: every (part of an) assessment is but a data-point/every data-point is optimised for learning by giving meaningful feedback to the learner/pass/fail decisions are not given on a single data-point

The rationale for these principles derives from the observation that ‘assessment drives learning behaviour’ and therefore a positive impact on learning approaches must be paramount. More adverse educational impacts are seen in typical modular, summative assessment systems (Al Kadri et al. 2009). How assessment drives learning is complex; however, it is becoming clear that both the (assessment) task and the assessment system design are important mechanisms, which are mediated by learner factors, such as the learner’s appraisal of the impact, perceived agency, and interpersonal factors (Cilliers et al. 2012a, 2012b; Schut et al. 2018). In addition, it has been found that feedback can be ignored in some traditional assessment systems (Harrison et al. 2013, 2015). These findings were paramount in reinforcing the objective of programmatic assessment to have assessment drive learning in a meaningful way and foster desirable learning approaches. The assessment programme is designed to optimise the learning function of assessment by the generation of meaningful, often narrative feedback and single assessments not being used for pass–fail decisions.

There was overall agreement and the majority of the 15 programmes that completed the inventory implemented these principles (Table 4). The need to generate meaningful feedback for learners was recognised as an important component. Feedback for complex skills is enhanced by narrative information (Govaerts and van der Vleuten 2013). Narrative feedback can also add meaning to standardised assessment (Tekian et al. 2017). In addition, the longitudinal organisation of learning and assessment curricular structures was mentioned both to enable feed-forward and to support longitudinal monitoring and guidance for learning. This also highlighted the importance of being conscious of the design of these longitudinal assessment curricular structures. For this, mapping or blueprinting of assessment was also indicated as an important component, which links to principle 4/5. It was also indicated that a change in mindset and assessment culture is needed. Indeed a discrepancy between a low-stakes design to stimulate learning and a high-stakes, summative perception of learners has been shown (Bok et al. 2013; Heeneman

et al. 2015). A deliberate design, with opportunities for the learners’ agency, a supportive assessment and/or feedback literacy programme for learners may help actualise the learning function of programmatic assessment (Price et al. 2012; Schut et al. 2018, 2020; Sutton 2012).

Principle 4/5: there is a mix of methods of assessment/the choice of method depends on the educational justification for using that method

An important rationale for these principles is that any assessment method has its limitations in terms of validity and reliability, and can be used for only one level of Miller’s pyramid (van der Vleuten et al. 2010). Therefore, an elaborate and purposeful mix of methods needs to be used to cover the whole pyramid and to ensure an appropriate mix of reliability and validity. In addition, the choice of any assessment format needs to be based on constructive alignment with the intended learning outcome and the teaching activities (Biggs 1996).

There was overall agreement with these principles and the majority of the 15 programmes that completed the inventory implemented them to some degree (Table 4). Some experts indicated a ‘partial’ importance, as this principle would be necessary in any educational and assessment design, not just in programmatic assessment. The components needed to apply these principles would be a deliberate choice of assessment methods guided by the principles of constructive alignment adhering to a blueprint. Guidelines of the blueprinting of courses have been described (Mookherjee et al. 2013; Villarroel et al. 2018), however in programmatic assessment, these blueprints need to cover the whole assessment design of the programme (Wilkinson and Tweed 2018), and governance and support by senior leadership and management is indispensable. In addition, the utility model was indicated as an important underlying concept of these principles (van der Vleuten 1996). This model characterises assessment utility by conceptually multiplying a number of elements on which assessment methods or instruments can be judged, such as reliability, validity, and educational impact. This conceptual multiplication model emphasises that if any element is zero, then the utility is zero. The experts indicated that in programmatic assessment, any assessment method can be used and be of value for the utility, but this can only be judged when seen within the context of the entire assessment programme.

Principle 6/7: the distinction between summative and formative is replaced by a continuum of stakes/decision-making on learner progress is proportionally related to the stakes

In programmatic assessment, the stakes of the assessment are conceptualised as a continuum from low- to high-stakes. This contrasts with the more traditional and binary dichotomy of formative versus summative assessment. In a low-stakes assessment, the results have no or limited consequences for the learner in terms of passing or failing, this datapoint instead is optimised for learning, as exemplified in principle 1/2/3. The high-stakes assessment or high-stakes decision, has important consequences, such as

graduation or promotion. The information from many low-stakes assessments contributes to the high-stakes decision, and the higher the stakes, proportionally more data points are needed for the decision (van der Vleuten et al. 2012).

There was overall agreement with these principles. However, several of the 15 programmes that completed the inventory indicated a mixed agreement about the degree of importance and whether it was easy to adhere to (Table 4). It was mentioned that a low-stakes assessment would still cause anxiety among learners, and it may not be easy for teachers to shift from a formative-summative paradigm to a low-high stakes continuum, as also indicated for principle 1/2/3. Regarding the perceptions of teachers, research has also shown that the use of programmatic assessment can positively transform teachers practices and assessment beliefs. Given principle 6/7, teachers can focus on the learning outcome of assessment (principle 1/2/3) and not the decision making outcome (see principle 9). This shift in teachers' focus was shown to reduce role conflicts, although the tension between teachers taking control and allowing learners' independence still needs careful navigation (Jamieson et al. 2021; Schut et al. 2020). Almost all programmes implemented principles 6/7 in their education and assessment program, using various formats, e.g. entrustable professional activities (ten Cate 2005; ten Cate and Scheele 2007), a high-stakes decision based on a comprehensive end of year portfolio assessment (Friedman Ben David et al. 2001; Tochel et al. 2009; van Tartwijk and Driessen 2009), and assessment of learning plans based on in-training assessment reports (Dawson et al. 2015; Laughlin et al. 2012), for which 'competence committees' were installed (see principle 9).

In the discussion with experts and Ottawa attendees prior to, and during the workshop, several points of attention were raised for the use of these principles. One point concerned the need for data saturation for high-stakes decisions. There is some evidence that consensus amongst decision makers is independent of the number of datapoints exceeding the required minimum, suggesting that data saturation can be obtained in a given context, with a defined minimum of datapoints (de Jong et al. 2019). Another point was raised on the need for psychological safety in teaching and working environments. There is little research yet on psychological safety in a setting where programmatic assessment is implemented. Learners can perceive low-stakes assessment as high-stakes and feel anxious. Tsuei et al. (2019) suggested that a number of features that learners would perceive as beneficial for feeling psychologically or educationally safe, are recognisable in the principles of programmatic assessment, such as having supportive relationships with peers and mentors and a focus on learning without considering consequences. Nevertheless, educational safety as a relational construct needs attention and awareness in any education design. Finally, the need was expressed to keep a focus on learner development and enable reflection, in the context of the high-stakes decision function of programmatic assessment. Reflection and self-monitoring have been recognised as important for professional development and performance, yet an overt instrumental and mandatory approach can lead to meaningless activities for the learners (Murdoch-Eaton and Sandars 2014). It has been shown that what

learners document on competency development in a portfolio can be influenced by tensions between learning and assessment, and the learners' perceptions about the purpose of the portfolio (Oudkerk Pool et al. 2020). However, learners also perceive the embedding of reflection or self-assessment in the learning function of programmatic assessment (principle 1/2/3) and the guidance from a coach (principle 11) are helpful for their learning (Heeneman et al. 2015).

Principle 8: assessment information triangulation across data-points, towards an appropriate framework

The principle of triangulation is based on domain-specificity; constructs such as competencies generalise well over assessment formats when the content domain is the same. This also opens up the possibility of making evidence-based decisions by attribute rather than by test format – for example determining if a learner has reached the required standard on history taking might draw on the history taking components of an OSCE, alongside the history taking components of a mini-CEX, and alongside the history taking components of a patient opinion survey. We see this triangulation of data in informing decisions as an important component to robust decision making (Norman et al. 1996; Schuwirth and Van Der Vleuten 2019). Thus, in programmatic assessment, assessment information that pertains to the same content is triangulated, to constructs such as knowledge, skills, and attitude or competencies. It will depend on the design, and national or legislative boundaries of the programme what is considered as an appropriate framework. In medical education, competencies are often used (Frank et al. 2010).

There was overall agreement with the principle, although some of the 15 programmes that completed the inventory (Table 4) indicated that it was less easy to adhere to, due to the need for a deliberate design, some form of technology to manage the data, and an understanding and support of this concept by faculty and the programme. The concept of triangulation can be difficult to translate into educational practice, as it often asks for a combination of numerical and narrative data. The end result is not a calculation but an informative narrative about and for the learner. This requires central governance of the educational and assessment design, alignment, faculty development, a necessary level of staff assessment literacy and expertise (Prentice et al. 2020; Schuwirth and Van Der Vleuten 2019), and establishment of effective group decision making processes which take a holistic view of the data (see later).

The inventory amongst the 15 programmes (Table 4) showed that most have implemented this principle with the components being: a careful design of educational activities, assessments, and assessment instruments, as well as high quality data aggregation in an appropriate manner. A robust system to collect all assessment and feedback information is essential (van der Vleuten et al. 2015). A technology supported approach, e.g. an electronic portfolio, is often used and could serve the purposes needed for programmatic assessment, (1) as a depository for all information (feedback forms, assessment results, minutes), (2) to facilitate administrative purposes of the programme of assessment (e.g. direct online completion of forms, such

as multisource feedback tools, loading of assessment and feedback forms via multiple platforms, managing access), (3) to support the triangulation function by generating overviews of aggregated datapoints using the (appropriate) framework, and (4) to support learners' self-assessment and agency (Tillema 2001; van Tartwijk and Driessen 2009). The technology approach chosen to collect the assessment and feedback information can, together with a coach (principle 11), support the learning function (principle 1/2/3), and the decision function of programmatic assessment (principle 9).

Principle 9: high-stakes decisions made in a credible and transparent manner, using a holistic approach

As embedded in principle 6/7, the high-stakes decision in programmatic assessment is based on many datapoints, on rich information originating from a broad sampling, across contexts, assessment methods, and diverse assessors (van der Vleuten et al. 2012, 2015). Given the high-stakes and prominent consequences, the procedures need to be trustworthy and credible. Procedural measures could include: appointment of an assessment committee of experts that are trained and can use narrative standards, rubrics or milestones; the provision of a justification for the decision; member-checking procedures, of the coach/mentor and the learner; instatement of appeal procedures. As expressed by van der Vleuten et al.: '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' (p. 643) (van der Vleuten et al. 2015).

Although there was general agreement on this principle and majority of the 15 programmes that completed the inventory (Table 4) have implemented it as such, it was also perceived by some as not easy to adhere to, due to the resources needed for these assessment procedures and/or committees, required leadership for acceptance of decisions by an expert group or committee and a mandate from institutional policies to enable enactment of these assessment procedures.

Many programmes used a group of experts to make the high-stakes decisions, e.g. clinical competency committees (Duitsman et al. 2019; Kinnear et al. 2018) or independent portfolio committees taking the view of mentor and learner into account with a member checking procedure (Driessen et al. 2012). The principles of group-decision making were emphasised, including the use of aggregated data to make an holistic decision, the importance of having a shared mental model and a proper method for sharing information (Hauer et al. 2016). In addition, the panel needs to be attuned to possible sources of bias associated with group decision making (Tweed and Wilkinson 2019). Approaches using mosaics of performance data and use of Bayesian networks have been proposed to support the committees in managing and maintaining overview of accumulating feedback and performance data, and informing the decision making (Pearce et al. 2021; Zoanetti and Pearce 2021). The need for credibility and transparency is not unique to programmatic assessment, all assessment procedures and formats need this. It is however important to realise that in programmatic assessment a holistic decision is made,

based on aggregated data that is presented in a variety of formats, meaning traditional grading rules or psychometrics are unlikely to be as applicable in the decision process.

Principle 10: intermediate review is made to discuss and decide with the learner on their progress

Given that the high-stakes decision at the end of a period, year or programme has substantial consequences, this must not come as a surprise for the learner (van der Vleuten et al. 2015). Therefore, it is imperative that the learner receives intermediate feedback on the potential decision and can act to improve if needed. This intermediate review can also be seen as an important procedural measure for ensuring the credibility of the high-stakes decision (see principle 9) (van der Vleuten et al. 2015). The intermediate review is based on fewer datapoints (proportionality, see principle 6/7) and is designed to give a 'diagnostic' message, how is the learner doing and what can be done. For this intermediate review, it is important that the learner is guided by a coach/mentor (principle 11), and that a feedback dialogue is in place. It is well known that feedback is most effective, when it is a 'loop', a cyclical process, involving a dialogue (Boud and Molloy 2012; Carless et al. 2011). The emphasis on the discussion and dialogue and the ability of the learner to act were also the rationales for the rephrasing of this principle after the workshop at the Ottawa conference.

Although there was agreement and most of the 15 programmes that completed the inventory (Table 4) have implemented, for some it was less easy to adhere to, because of the necessary resources and the need to explicitly incorporate an intermediate moment in the design of the curriculum. Often the intermediate moment was implemented as a formal moment in time halfway through a period or year, integrated as part of the process of mentor meetings, or done by the supervisors. It was also indicated that the presence of an intermediate review signified that the programme takes care of the learner, in facilitating the learning. The role of the learners themselves in using the feedback, and follow-up of feedback was seen as very important.

Principle 11: learners have recurrent learning meetings with (faculty) mentors/coaches using a self-analysis of all assessment data

As indicated above (principles 1/2/3 and 10), feedback is essential for learning and professional development. The use of that feedback by the learners is often scaffolded in self-analysis or reflection (Sargeant et al. 2009). Learners do not appreciate reflective activities as more than tick-box exercises (de la Croix and Veen 2018); however, they do see the value of reflection as part of a dialogue with a mentor (Driessen et al. 2012; Heeneman et al. 2015). It is well known that self-direction and reflection require direction and guidance by a mentor or coach (Knowles 1975; Pilling-Cormick 1997). Therefore, this guidance by a mentor is an important principle in programmatic assessment (van der Vleuten et al. 2012, 2015).

There was overall agreement and the majority of the 15 programmes that completed the inventory (Table 4)

implemented this principle. Lack of resources and lack of (trained) staff were factors that made this principle less easy to adhere to. Most programmes used dedicated staff mentors/coaches, or in post-graduate training the programme director was involved. It was clear that the size of the programme also mattered; if many learners were present, and resources limited, the choice could be made to have no mentoring system or a limited number of contacts throughout the year. Points of attention were the importance of a faculty development programme and to be aware of potential tensions that mentors or coaches might perceive when the portfolio that serves as a guidance instrument (see principle 8) in the mentoring relationship, is also used in the high-stakes decision making (Anderson and DeMeulle 1998; Castanelli et al. 2020; Heeneman and de Grave 2017).

Principle 12: programmatic assessment seeks to gradually increase the learner's agency and accountability for their own learning through the learning being tailored to support individual learning priorities

For the learning function of assessment (principle 1/2/3), assessment and feedback are designed as low-stakes, and the continuous flow of information fosters self-regulated learning. Frameworks such as the self-determination theory and self-regulated learning indeed support the importance of learners' motivation and agency for learning (Panadero 2017; Zimmerman 1989). Schut et al. identified that, in the context of programmatic assessment, the feeling of being in control, or agency, was essential for the learners' perception of assessment stakes (Schut et al. 2018). Programme features were an important factor in whether learners were able to take control over the assessment and perceive it as low-stakes, i.e. a sense of agency was encouraged when the programme allowed the learner to initiate their own assessment or select the evidence for their progress (Schut et al. 2018).

This was perceived as the most complex principle and rephrasing was needed to convey the message and implications. After rephrasing there was overall agreement with the principle, although it was not easy to adhere to, and implementation was partial in some of the programmes that completed the inventory (Table 4). It was indicated that agency and accountability are important for all learners, both for those that do well and those that struggle. This is challenging as for the learners that struggle, coaches and staff are more likely to step in and take action (Heeneman and de Grave 2017), and remediation is controlled and regulated by staff (Ellaway et al. 2018). Yet the focus on learning in programmatic assessment suggests those already doing well are supported and encouraged to do even better, reinforcing the importance of lifelong learning for all health care practitioners.

Table 2 presents the final Ottawa 2020 consensus principles of programmatic assessment. The principles in Table 2 are not to be considered as items of a checklist that need to be fulfilled in order to call the programme of assessment, programmatic. As indicated earlier, the principles represent a conceptual view on education, assessment and its alignment. Programmatic assessment is not a recipe and may

have many different manifestations. These manifestations may nevertheless be considered as programmes in which programmatic assessment is leading the educational design and maximises the learning and decision function of assessment in that context. Some of these manifestations will be elaborated on in part 2, the practical implementation of the consensus principles (Torre et al. 2021).

Recommendations for future work

The work and proceedings for this Ottawa 2020 consensus statement (part 1) on the principles of programmatic assessment led to a number of important insights. First, a significant aspect in the programmatic assessment model is the interlinking of certain principles, e.g. for the intermediate progress meeting (principle 10), guidance by a mentor is needed (principle 11). The finding that the principles depend on each other in practice, is important to take into account in the choice for programmatic assessment as central to the assessment and educational design of a curriculum. An important question is whether there are a certain number of principles, or whether there are specific principles that are needed in the design to lead to the desired impact on the learning and decision function of assessment. In other words, are there principles without which a system could not be called programmatic and/or are there a certain number of principles that need to be applied, before a programmatic approach is realised? Here, the comparison to the implementation of other educational formats across contexts may be useful, e.g. problem-based learning (PBL). PBL can have many manifestations or hybrid approaches, as a result of compromises on the original intended model. Studies have shown that the outcome of PBL may then give a 'hybrid' success (Frambach et al. 2012), and also in programmatic assessment, the partial implementation of certain principles may give unwanted side effects, e.g. low-stakes assessment that is not perceived as such by the learners (Bok et al. 2013; Heeneman et al. 2015; Schut et al. 2018). An important question is whether the implementation itself was not optimal and therefore led to a hybrid outcome, or was a certain principle only partially implemented, e.g. a number of assessments are present that yield individual summative decisions, and that led to the hybrid outcome. However, there could also be principles that are easier to implement or address a certain need in a particular organisation or context, which might create effective hybrids. There are examples of such hybrid or incremental approaches (Bala et al. 2020; Bate et al. 2020) and McDowell reported on the implementation of programme focused assessment in a number of Higher Education institutes in the United Kingdom, which featured horizontal integrative assessment across stages/years of a programme (McDowell 2012). More research on the effects of various implementations of programmatic assessment on outcomes such as learning behaviour and decision making is definitely needed. It is however emphasised and considered as a key feature of programmatic assessment that principles pertaining to both the learning function (i.e. principle 1/2/3/11/12) and the decision function of assessment (i.e. principle 7/8/9/10) should be present if programmatic assessment is central to the assessment and educational design of a curriculum.

Different manifestations of programmatic assessment were seen in the actual practices of the experts' programmes that contributed to this consensus statement, and this should be encouraged as innovation can arise from diversity. Further research may better understand which individual principles, or the interaction therein, have the greatest impact on the desired outcomes, including the competencies of health care professionals society desires, and where possible, impact on patient and system outcomes.

Second, the principles on the continuum of stakes (principle 6/7) and learner agency (principle 12) gave the most varied responses in terms of importance and adherence (Table 4). Indeed the continuum of stakes is an important theoretical foundation of the programmatic assessment model, but the dimensions of this continuum were considered difficult to grasp and adhere to. For example: When is a low-stakes assessment 'truly' low-stakes? How should we come to a meaningful and reliable high-stakes decision at the level of the intended learning outcomes? And how can the full continuum be employed? Although studies have shed light on the perceptions of both learners and teachers on low-stakes assessments (Bok et al. 2013; Dart et al. 2021; Heeneman et al. 2015; Schut et al. 2018, 2020) and the trustworthiness of the high-stakes decision (de Jong et al. 2019), and suggestions were made to extrapolate from clinical decision-making and jury decision-making for learners progress decisions (Tweed and Wilkinson 2019), more work needs to be done on the use and meaning of the full range of the low- to high-stakes continuum.

Regarding principle 12, learner agency was perceived as important but also difficult to achieve. It was clear that this would need a change in both the curriculum and assessment design and even more important the alignment (Kulasegaram et al. 2018). As shown by Watling et al., 'agency is (hard) work', learners may need to resist social and professional expectations, and support/coaching was deemed as fruitful and needed (principle 11) (Watling et al. 2021). Institutional policies and accreditation bodies can create tensions by emphasising the need to attain and safeguard the intended learning outcomes of a programme versus the desired autonomy or agency of the learner to maximise self-regulation and self-determination of learning. In addition, the teachers' role was shown to be important: when teachers are more focused on the conception of accounting and control, this could lead to tensions. Teachers may struggle between being in control and permitting learner agency (Schut et al. 2020). The tensions caused by either teachers or legislative bodies are difficult to navigate, and need further exploration and research. Faculty development and communities of practice for teachers is crucial.

Third, it is clear that the context is a very significant influence on the implementation and the potential outcomes of programmatic assessment. It is important to note that the experts of this consensus statement were based in European, North-American, and Australasia regions. It is well known that cultural aspects influence assessment beliefs and systems (Wong 2011). Calls for a change of assessment practices in other regions are made (Khan 2018) and more studies on programmatic assessment in other regions or cultures are highly recommended.

Conclusions

In conclusion, we present 12 principles that are considered as important and recognisable facets of programmatic assessment. It is important to note that this consensus is based on current insights and practices. Follow-up research and implementation practices may lead to amendments or change in the consensus.

An inventory amongst experts and their programmes showed that these principles were used and implemented, albeit with a range of approaches and rigor, suggesting that programmatic assessment is a realistic assessment model that can be implemented. The variability was related to various context factors such as programme size, institutional barriers, legislation restrictions, available resources, level of assessment literacy and underlying attitudes to change. Sharing knowledge of how programmatic assessment is being operationalised in different contexts may help educators in signifying their current or future plans for the implementation journey of programmatic assessment in their programmes. Such a journey is never done, requiring deliberate and ongoing attention to contextual, system, teacher, and learner aspects that ultimately interact to allow programmatic assessment to fully leverage, on a sustained basis, its learning and decision function. This is further elaborated on in a follow-up data collection in 15 programmes in part 2, the practical implementation of the Ottawa 2020 consensus principles (part 2) (Torre et al. 2021).

Acknowledgements

The input of the expert group and attendees of the Ottawa 2020 conference is greatly appreciated. The expert group consisted of the following esteemed colleagues: Lambert Schuwirth (Adelaide, Australia), Harold Bok (Utrecht, the Netherlands), Beth Bierer (Cleveland, USA), Teresa Chan (McMaster, Canada), Luke Dawson (Liverpool, UK), Lubberta de Jong (Utrecht, the Netherlands), Paul Dilena and Jill Bendon (Unley SA, Australia), Marjan Govaerts (Maastricht, the Netherlands), Sylvia Heeneman (Maastricht, the Netherlands), Jaime Jamieson (Perth, Australia), Tom Laughlin (Moncton, Canada), Neil Rice (Exeter, UK), Anna Ryan (Melbourne, Australia), Suzanne Schut (Maastricht, the Netherlands), Glendon Tait (Toronto, Canada), Cees van der Vleuten (Maastricht, The Netherlands), Kiran Veerapen (Vancouver, Canada), Tim Wilkinson (Otago, New Zealand), Dario Torre (Bethesda, USA), Adrian Freeman (Exeter, UK).

Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

Notes on contributors

Sylvia Heeneman, PhD, Professor of Health Profession Education at the School of Health Profession Education, Department of Pathology, Faculty of Health, Medicine and Life Sciences, Maastricht University, the Netherlands.

Lubberta H. de Jong, MSc DVM, PhD candidate at the Centre of Quality Improvement in Veterinary Education, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands.

Luke J. Dawson, BSc BDS PhD FDSRCS(Eng) FHEA MA (TLHE) NTF, Professor of Dental Education, Director of Undergraduate Education, School of Dentistry, Liverpool, UK.

Tim J. Wilkinson, MB ChB MD PhD(Otago) M Clin Ed(UNSW) FRACP FRCP(London) FANZAHPE, Director of the University of Otago MBChB programme, Education unit, University of Otago, Christchurch, New Zealand.

Anna Ryan, MBBS, PhD, Associate Professor, Director of Assessment, Department of Medical Education, Melbourne Medical School, University of Melbourne, Australia.

Glendon R. Tait, MD, MSc, Director of Student Assessment, MD Program at the University of Toronto; Associate Professor Department of Psychiatry and practices Consultation-Liaison Psychiatry with Sinai Health System, and The Wilson Centre, University of Toronto, Canada.

Neil Rice, MSc, Head of Psychometrics and Informatics, University of Exeter, College of Medicine and Health, Exeter, UK.

Dario Torre, MD, MPH, PhD, Professor of Medicine and Associate Director for Program Evaluation and Long Term Outcomes at Uniformed Services University of Health Sciences in Bethesda, Maryland, USA.

Adrian Freeman, MD, MMedSci, FRCGP, FAcadMed, Professor of Medical Education at University of Exeter Medical School, examiner for the Royal College of General Practitioners, President of the European Board of Medical Assessors and Deputy chair of the GMC Panel for Tests of Competence, University of Exeter, College of Medicine and Health, Exeter, UK.

Cees P. M. van der Vleuten, PhD, Professor of Education, Faculty of Health Medicine and Life Sciences, Maastricht University, The Netherlands.

ORCID

Tim J. Wilkinson  <http://orcid.org/0000-0002-4080-4164>

Neil Rice  <http://orcid.org/0000-0002-9718-1316>

Dario Torre  <http://orcid.org/0000-0002-4924-4888>

Cees P. M. van der Vleuten  <http://orcid.org/0000-0001-6802-3119>

References

- Al Kadri H, Al-Moamary M, Van Der Vleuten C. 2009. Students' and teachers' perceptions of clinical assessment program: a qualitative study in a PBL curriculum. *BMC Res Notes*. 2(1):263.
- Anderson R, Demeulle L. 1998. Portfolio use in twenty-four teacher education programs. *Teach Educ Quart*. 25:23–31.
- Bala L, Van Der Vleuten C, Freeman A, Torre D, Heeneman S, Sam A. 2020. COVID-19 and programmatic assessment. *Clin Teach*. 17(4):420–422.
- Bate F, Fyfe S, Griffiths D, Russell K, Skinner C, Tor E. 2020. Does an incremental approach to implementing programmatic assessment work? Reflections on the change process. *MedEdPublish*. 9(1).
- Biggs J. 1996. Enhancing teaching through constructive alignment. *High Educ*. 32(3):347–364.
- Bok H, Teunissen P, Favier RP, Rietbroek N, Theyse L, Brommer H, Haarhuis J, Van Beukelen P, Van Der Vleuten C, Jaarsma D. 2013. Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Med Educ*. 13:123.
- Boud D, Molloy E. 2012. Rethinking models of feedback for learning: the challenge of design. *Assess Eval High Educ*. 2012:1–15.
- Carless D, Salter D, Yang M, Lam J. 2011. Developing sustainable feedback practices. *Stud High Educ*. 36(4):395–407.
- Castanelli D, Weller J, Molloy E, Bearman M. 2020. Shadow systems in assessment: how supervisors make progress decisions in practice. *Adv Health Sci Educ*. 25(1):131–147.
- Chan T, Sherbino J. 2015. The McMaster Modular Assessment Program (McMAP): a theoretically grounded work-based assessment system for an emergency medicine residency program. *Acad Med*. 90(7):900–905.
- Cilliers F, Schuwirth L, Herman N, Adendorff H, Van Der Vleuten C. 2012a. A model of the pre-assessment learning effects of summative assessment in medical education. *Adv Health Sci Educ*. 17(1):39–53.
- Cilliers F, Schuwirth L, Van Der Vleuten C. 2012b. Modelling the pre-assessment learning effects of assessment: evidence in the validity chain. *Med Educ*. 46(11):1087–1098.
- Dannefer EF, Henson L. 2007. The portfolio approach to competency-based assessment at the Cleveland Clinic Lerner College of Medicine. *Acad Med*. 82:493–502.
- Dannefer EF, Prayson RA. 2013. Supporting students in self-regulation: use of formative feedback and portfolios in a problem-based learning setting. *Med Teach*. 35:1–6.
- Dart J, Twohig C, Anderson A, Bryce A, Collins J, Gibson S, Kleve S, Porter J, Volders E, Palermo C. 2021. The value of programmatic assessment in supporting educators and students to succeed: a qualitative evaluation. *J Acad Nutr Diet*. DOI:10.1016/j.jand.2021.01.013.
- Dawson L, Mason B, Balmer C, Jimmieson P. 2015. Developing professional competence using integrated technology-supported approaches: a case study in dentistry. In: Fry H, Ketteridge S, Marshall S, editors. *A handbook for teaching and learning in higher education enhancing academic practice*. London; New York: Routledge.
- De Jong L, Bok H, Kremer W, Van Der Vleuten C. 2019. Programmatic assessment: can we provide evidence for saturation of information? *Med Teach*. 41(6):678–682.
- De La Croix A, Veen M. 2018. The reflective zombie: problematizing the conceptual framework of reflection in medical education. *Perspect Med Educ*. 7(6):394–400.
- Driessen E, Van Tartwijk J, Govaerts M, Teunissen P, Van Der Vleuten C. 2012. The use of programmatic assessment in the clinical workplace: a Maastricht case report. *Med Teach*. 34(3):226–231.
- Duitsman ME, Fluit CRMG, van Alfen-van der Velden JAEM, de Visser M, Ten Kate-Booij M, Dolmans DHJM, Jaarsma DADC, de Graaf J. 2019. Design and evaluation of a clinical competency committee. *Perspect Med Educ*. 8(1):1–8.
- Ellaway R, Chou C, Kalet A. 2018. Situating remediation: accommodating success and failure in medical education systems. *Acad Med*. 93(3):391–398.
- Frambach J, Driessen E, Chan L, Van Der Vleuten C. 2012. Rethinking the globalisation of problem-based learning: how culture challenges self-directed learning. *Med Educ*. 46(8):738–747.
- Frank JR, Snell LS, Cate OT, Holmboe ES, Carraccio C, Swing SR, Harris P, Glasgow NJ, Campbell C, Dath D, et al. 2010. Competency-based medical education: theory to practice. *Med Teach*. 32(8):638–645.
- Freeman A, Ricketts C. 2010. Choosing and designing knowledge assessments: experience at a new medical school. *Med Teach*. 32(7):578–581.
- Friedman Ben David M, Davis MH, Harden RM, Howie PW, Ker J, Pippard MJ. 2001. AMEE Medical Education Guide No. 24: portfolios as a method of student assessment. *Med Teach*. 23(6):535–551.
- Govaerts M, Van Der Vleuten C. 2013. Validity in work-based assessment: expanding our horizons. *Med Educ*. 47(12):1164–1174.
- Harrison CJ, Konings KD, Molyneux A, Schuwirth L, Wass V, Van Der Vleuten C. 2013. Web-based feedback after summative assessment: how do students engage? *Med Educ*. 47(7):734–744.
- Harrison C, Könings K, Schuwirth L, Wass V, Van Der Vleuten C. 2015. Barriers to the uptake and use of feedback in the context of summative assessment. *Adv Health Sci Educ*. 20(1):229–245.
- Hauer K, Cate O, Boscardin C, Iobst W, Holmboe E, Chesluk B, Baron R, O'Sullivan P. 2016. Ensuring resident competence: a narrative review of the literature on group decision making to inform the work of clinical competency committees. *J Grad Med Educ*. 8(2):156–164.
- Heeneman S, De Grave W. 2017. Tensions in mentoring medical students toward self-directed and reflective learning in a longitudinal portfolio-based mentoring system – an activity theory analysis. *Med Teach*. 39(4):368–376.
- Heeneman S, Oudkerk Pool A, Schuwirth L, Van Der Vleuten C, Driessen E. 2015. The impact of programmatic assessment on student learning: theory versus practice. *Med Educ*. 49(5):487–498.
- Jamieson J, Hay M, Gibson S, Palermo C. 2021. Implementing programmatic assessment transforms supervisor attitudes: an explanatory sequential mixed methods study. *Med Teach*. 43:1–9.
- Jamieson J, Jenkins G, Beatty S, Palermo C. 2017. Designing programmes of assessment: a participatory approach. *Med Teach*. 39(11):1182–1188.

- Khan R. 2018. Measuring learning of medical students through 'programmatic assessment'. *Pak J Med Sci.* 34:3–4.
- Kinnear B, Warm E, Hauer K. 2018. Twelve tips to maximize the value of a clinical competency committee in postgraduate medical education. *Med Teach.* 40(11):1110–1115.
- Knowles M. 1975. *A guide for learners and teachers.* England Cliffs: Prentice Hall/Cambridge.
- Kulasegaram K, Mylopoulos M, Tonin P, Bernstein S, Bryden P, Law M, Lazor J, Pittini R, Sockalingam S, Tait G, et al. 2018. The alignment imperative in curriculum renewal. *Med Teach.* 40(5):443–448.
- Laughlin T, Brennan A, Brailovsky C. 2012. Effect of field notes on confidence and perceived competence: survey of faculty and residents. *Can Fam Phys.* 58:e352–e356.
- Lingard L. 2007. The rhetorical 'turn' in medical education: what have we learned and where are we going? *Adv Health Sci Educ Theory Pract.* 12(2):121–133.
- Lingard L. 2009. What we see and don't see when we look at 'competence': notes on a god term. *Adv Health Sci Educ.* 14(5):625–628.
- McDowell L. 2012. Programme focused assessment – a short guide. Northumbria University. <https://www.bradford.ac.uk/pass/resources/short-guide.pdf>.
- Mookherjee S, Chang A, Boscardin C, Hauer K. 2013. How to develop a competency-based examination blueprint for longitudinal standardized patient clinical skills assessments. *Med Teach.* 35(11):883–890.
- Murdoch-Eaton D, Sandars J. 2014. Reflection: moving from a mandatory ritual to meaningful professional development. *Arch Dis Child.* 99(3):279–283.
- Norcini J, Anderson B, Bollela V, Burch V, Costa MJ, Duvivier R, Galbraith R, Hays R, Kent A, Perrott V, et al. 2011. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. *Med Teach.* 33(3):206–214.
- Norcini J, Anderson M, Bollela V, Burch V, Costa M, Duvivier R, Hays R, Palacios Mackay M, Roberts T, Swanson D. 2018. 2018 consensus framework for good assessment. *Med Teach.* 40(11):1102–1109.
- Norman G, Swanson D, Case SM. 1996. Conceptual and methodological issues in studies comparing assessment formats. *Teach Learn Med.* 8(4):208–216.
- Pool AO, Jaarsma A, Driessen E, Govaerts M. 2020. Student perspectives on competency-based portfolios: does a portfolio reflect their competence development? *Perspect Med Educ.* 9(3):166–172.
- Panadero E. 2017. A review of self-regulated learning: six models and four directions for research. *Front Psychol.* 8:422.
- Pearce J, Reid K, Chiavaroli N, Hyam D. 2021. Incorporating aspects of programmatic assessment into examinations: aggregating rich information to inform decision-making. *Med Teach.* 43:1–8.
- Pilling-Cormick J. 1997. Transformative and self-directed learning in practice. *New Dir Adult Cont Educ.* 1997(74):69–77.
- Prentice S, Benson J, Kirkpatrick E, Schuwirth L. 2020. Workplace-based assessments in postgraduate medical education – a hermeneutic review. *Med Educ.* 54(11):981–992.
- Price M, Rust C, O'Donovan B, Handley K, Bryant R. 2012. *Assessment literacy: the foundation for improving student learning.* Oxford: Oxford Brookes University.
- Ryan A, McColl G, O'Brien R, Chiavaroli N, Judd T, Finch S, Swanson D. 2017. Tensions in post-examination feedback: information for learning versus potential for harm. *Med Educ.* 51(9):963–973.
- Sargeant J, Mann K, Van Der Vleuten C, Metsemakers J. 2009. Reflection: a link between receiving and using assessment feedback. *Adv Health Sci Educ.* 14(3):399–410.
- Schut S, Driessen E, Van Tartwijk J, Van Der Vleuten C, Heeneman S. 2018. Stakes in the eye of the beholder: an international study of learners' perceptions within programmatic assessment. *Med Educ.* 52(6):654–663.
- Schut S, Heeneman S, Bierer B, Driessen E, Van Tartwijk J, Van Der Vleuten C. 2020. Between trust and control: teachers' assessment conceptualisations within programmatic assessment. *Med Educ.* 54(6):528–537.
- Schut S, Maggio L, Heeneman S, Van Tartwijk J, Van Der Vleuten C, Driessen E. 2021. Where the rubber meets the road – an integrative review of programmatic assessment in health care professions education. *Perspect Med Educ.* 10(1):6–13.
- Schuwirth L, Valentine N, Dilena P. 2017. An application of programmatic assessment for learning (PAL) system for general practice training. *GMS J Med Educ.* 34:Doc56.
- Schuwirth L, Van Der Vleuten C. 2011. Programmatic assessment: from assessment of learning to assessment for learning. *Med Teach.* 33(6):478–485.
- Schuwirth L, Van Der Vleuten C. 2019. Current assessment in medical education: programmatic assessment. *J Appl Test Technol.* 20:2–10.
- Sutton P. 2012. Conceptualizing feedback literacy: knowing, being, and acting. *Inn Educ Teach Int.* 49(1):31–40.
- Tekian A, Watling C, Roberts T, Steinert Y, Norcini J. 2017. Qualitative and quantitative feedback in the context of competency-based education. *Med Teach.* 39(12):1245–1249.
- Ten Cate O. 2005. Entrustability of professional activities and competency-based training. *Med Educ.* 39(12):1176–1177.
- Ten Cate O, Scheele F. 2007. Viewpoint: competency-based postgraduate training: can we bridge the gap between theory and clinical practice? *Acad Med.* 82:542–547.
- Tillema H. 2001. Portfolios as developmental assessment tools. *Int J Train Dev.* 5(2):126–135.
- Tochel C, Haig A, Hesketh A, Cadzow A, Beggs K, Colthart I, Peacock H. 2009. The effectiveness of portfolios for post-graduate assessment and education: BEME Guide No 12. *Med Teach.* 31(4):299–318.
- Torre D, Rice N, Ryan A, Bok H, Dawson L, Bierer B, Wilkinson T, Tait G, Laughlin T, Veerapen K, et al. 2021. Ottawa 2020 consensus statements for programmatic assessment 2: implementation and practice. *Med Teach.* in press.
- Tsuei S, Lee D, Ho C, Regehr G, Nimmon L. 2019. Exploring the construct of psychological safety in medical education. *Acad Med.* 94:528–535.
- Tweed M, Wilkinson T. 2019. Student progress decision-making in programmatic assessment: can we extrapolate from clinical decision-making and jury decision-making? *BMC Med Educ.* 19(1):176.
- Van Der Vleuten C. 1996. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ Theory Pract.* 1(1):41–67.
- Van Der Vleuten C, Heeneman S, Schut S. 2019. Programmatic assessment: an avenue to a different assessment culture. In: Yudkowsky R, Soo Park Y, Downing S, editors. *Assessment in health professions education.* New York: Routledge.
- Van Der Vleuten C, Heeneman S, Schuwirth L. 2017. Programmatic assessment. In: Dent J, Harden R, Hunt D, editors. *A practical guide for medical teachers.* Edinburgh: Elsevier.
- Van Der Vleuten C, Schuwirth L. 2005. Assessing professional competence: from methods to programmes. *Med Educ.* 39(3):309–317.
- Van Der Vleuten C, Schuwirth L, Driessen E, Dijkstra J, Tigelaar D, Baartman L, Van Tartwijk J. 2012. A model for programmatic assessment fit for purpose. *Med Teach.* 34(3):205–214.
- Van Der Vleuten C, Schuwirth L, Driessen E, Govaerts M, Heeneman S. 2015. Twelve tips for programmatic assessment. *Med Teach.* 37(7):641–646.
- Van Der Vleuten C, Schuwirth L, Scheele F, Driessen E, Hodges B. 2010. The assessment of professional competence: building blocks for theory development. *Best Pract Res Clin Obstet Gynaecol.* 24(6):703–719.
- Van Tartwijk J, Driessen E. 2009. Portfolios for assessment and learning: AMEE Guide no. 45. *Med Teach.* 31(9):790–801.
- Villarroel V, Bloxham S, Bruna D, Bruna C, Herrera-Seda C. 2018. Authentic assessment: creating a blueprint for course design. *Assess Eval High Educ.* 43(5):840–854.
- Watling C, Ginsburg S, Ladonna K, Lingard L, Field E. 2021. Going against the grain: an exploration of agency in medical learning. *Med Educ.*
- Wilkinson T, Tweed M. 2018. Deconstructing programmatic assessment. *Adv Med Educ Pract.* 9:191–197.
- Wilkinson T, Tweed M, Egan T, Ali A, Mckenzie J, Moore M, Rudland J. 2011. Joining the dots: conditional pass and programmatic assessment enhances recognition of problems with professionalism and factors hampering student progress. *BMC Med Educ.* 11(1):1–9.
- Wong A. 2011. Culture in medical education: comparing a Thai and a Canadian residency programme. *Med Educ.* 45(12):1209–1219.
- Zimmerman BJ. 1989. A social cognitive view of self-regulated academic learning. *J Educ Psychol.* 81(3):329–339.
- Zoanetti N, Pearce J. 2021. The potential use of Bayesian networks to support committee decisions in programmatic assessment. *Med Educ.* 55(7):808–817.