

Ottawa 2020 consensus statements for programmatic assessment 2: Implementation and practice

Citation for published version (APA):

Torre, D., Rice, N. E., Ryan, A., Bok, H., Dawson, L. J., Bierer, B., Wilkinson, T. J., Tait, G. R., Laughlin, T., Veerapen, K., Heeneman, S., Freeman, A., & van der Vleuten, C. (2021). Ottawa 2020 consensus statements for programmatic assessment 2: Implementation and practice. *Medical Teacher*, 43(10), 1149-1160. <https://doi.org/10.1080/0142159X.2021.1956681>

Document status and date:

Published: 03/10/2021

DOI:

[10.1080/0142159X.2021.1956681](https://doi.org/10.1080/0142159X.2021.1956681)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.




Ottawa 2020 consensus statements for programmatic assessment – 2. Implementation and practice

Dario Torre, Neil E. Rice, Anna Ryan, Harold Bok, Luke J. Dawson, Beth Bierer, Tim J. Wilkinson, Glendon R. Tait, Tom Laughlin, Kiran Veerapen, Sylvia Heeneman, Adrian Freeman & Cees van der Vleuten


To cite this article: Dario Torre, Neil E. Rice, Anna Ryan, Harold Bok, Luke J. Dawson, Beth Bierer, Tim J. Wilkinson, Glendon R. Tait, Tom Laughlin, Kiran Veerapen, Sylvia Heeneman, Adrian Freeman & Cees van der Vleuten (2021) Ottawa 2020 consensus statements for programmatic assessment – 2. Implementation and practice, *Medical Teacher*, 43:10, 1149-1160, DOI: [10.1080/0142159X.2021.1956681](https://doi.org/10.1080/0142159X.2021.1956681)

To link to this article: <https://doi.org/10.1080/0142159X.2021.1956681>

 View supplementary material [↗](#)

 Published online: 30 Jul 2021.

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

 Article views: 950






 View related articles [↗](#)

 View Crossmark data [↗](#)

 Citing articles: 2 View citing articles [↗](#)



Ottawa 2020 consensus statements for programmatic assessment – 2. Implementation and practice

Dario Torre^a , Neil E. Rice^b , Anna Ryan^c, Harold Bok^d, Luke J. Dawson^e, Beth Bierer^f ,
Tim J. Wilkinson^g , Glendon R. Tait^h, Tom Laughlinⁱ, Kiran Veerapen^j, Sylvia Heeneman^k, Adrian Freeman^b
and Cees van der Vleuten^k 

^aDepartment of Medicine, Uniformed Services University of Health Sciences, Bethesda, MD, USA; ^bCollege of Medicine and Health, University of Exeter Medical School, Exeter, UK; ^cDepartment of Medical Education, Melbourne Medical School, University of Melbourne, Melbourne, Australia; ^dDepartment of Population Health Sciences, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands; ^eSchool of Dentistry, University of Liverpool, Liverpool, UK; ^fCleveland Clinic Lerner College of Medicine of Case Western Reserve University, Cleveland, OH, USA; ^gEducation unit, University of Otago, Christchurch, New Zealand; ^hMD Program, Dept. of Psychiatry, and The Wilson Centre, University of Toronto, Toronto, Canada; ⁱDepartment of Family Medicine, Dalhousie University, Halifax, Canada; ^jFaculty of Medicine, University of British Columbia, Vancouver, Canada; ^kDepartment of Educational Development and Research, School of Health Profession Education, Maastricht University, Maastricht, The Netherlands

ABSTRACT

Introduction: Programmatic assessment is a longitudinal, developmental approach that fosters and harnesses the learning function of assessment. Yet the implementation, a critical step to translate theory into practice, can be challenging. As part of the Ottawa 2020 consensus statement on programmatic assessment, we sought to provide descriptions of the implementation of the 12 principles of programmatic assessment and to gain insight into enablers and barriers across different institutions and contexts.

Methods: After the 2020 Ottawa conference, we surveyed 15 Health Profession Education programmes from six different countries about the implementation of the 12 principles of programmatic assessment. Survey responses were analysed using a deductive thematic analysis.

Results and Discussion: A wide range of implementations were reported although the principles remained, for the most part, faithful to the original enunciation and rationale. Enablers included strong leadership support, ongoing faculty development, providing students with clear expectations about assessment, simultaneous curriculum renewal and organisational commitment to change. Most barriers were related to the need for a paradigm shift in the culture of assessment. Descriptions of implementations in relation to the theoretical principles, across multiple educational contexts, coupled with explanations of enablers and barriers, provided new insights and a clearer understanding of the strategic and operational considerations in the implementation of programmatic assessment. Future research is needed to further explore how contextual and cultural factors affect implementation.

KEYWORDS

Assessment; feedback; theory; general; portfolio

Introduction

At the 2020 Ottawa conference, consensus was reached on twelve theoretical principles of programmatic assessment (Heeneman et al. 2021). The consensus group, recognising the challenges of implementation (Norcini and Burch 2007; Norcini et al. 2018), identified that, in order for programmatic assessment and its principles to exert their educational benefits, it is important to share a common understanding of how these principles are implemented across different contexts and regulatory settings. As a follow up to the Ottawa 2020 consensus statement, it was therefore agreed to collect descriptions of implementations of programmatic assessment in order to identify enablers and barriers and to gain insight into those aspects that have been more readily implemented than others.

Programmatic assessment is a longitudinal, developmental approach that fosters and harnesses the learning function of assessment, using a mix of assessment methods for the purpose of providing feedback to learners and informing credible assessment decisions (Schuwirth and Van der

Practice points

- The descriptions of programmatic assessment implementations contribute to enhance our knowledge and understanding of how the principles of programmatic assessment were applied across different programmes in undergraduate and post-graduate health professions programmes.
- There was a range of implementations tailored and adapted to meet the needs of the local context.
- An individual and organisational shift from a traditional assessment paradigm is critical for a successful implementation of programmatic assessment.
- Implementation of programmatic assessment takes time, needs strong leadership and an equal buy-in from both teachers and learners.
- Implementing programmatic assessment provides insight into curriculum development, enhances feedback literacy and the agency of the learner, and improves the data available for decision making.

Vleuten 2011; Van Der Vleuten et al. 2012). It has been suggested that a programmatic assessment system may be more effective and valid for discerning true educational progress and performance (Cook et al. 2015). The validity of programmatic assessment has previously been appraised by mapping the overarching structure of Kane's validity theory to the components and tenets of programmatic assessment (Schuwirth and van der Vleuten 2012). Furthermore, there is an increasing body of evidence about the educational benefits and theoretical tenets of programmatic assessment (Van Der Vleuten and Schuwirth 2005; Bierer et al. 2015; Heeneman et al. 2015; Schut et al. 2021).

There are, however, inherent issues with change in any established educational system. The notion that context is likely to affect change and explain any differences in implementations has been described (Damschroder et al. 2009). Political, economic and social context, organisational culture, leadership engagement, characteristics of individuals including stakeholder's acceptance, ability and willingness to embrace change (Wilson and Sloane 2000; Damschroder et al. 2009) may all hinder or facilitate progress and affect change.

Globally, several programmes of education have successfully innovated with programmatic assessment to varying extents (Dannefer and Henson 2007; Freeman and Ricketts 2010; Driessen et al. 2012; Schut et al. 2018; De Jong et al. 2019; Sherbino et al. 2020). Nevertheless, there is limited understanding of implementation approaches taken across different contexts. Health Professions Education programmes involve complex, dynamic, multilevel systems incorporating a number of interacting factors that ultimately affect implementation. Learning how programmatic assessment has been implemented and incorporated in different educational contexts is essential to gain insight on how the stated principles have been translated into educational practices, whilst identifying potential enablers and barriers that may impact implementation processes. These insights can be valuable for educators and institutions that currently use programmatic assessment and for those who are planning to implement it.

In this paper we report on descriptions of implementation of programmatic assessment and how different descriptions align with and adhere to the principles established in the consensus statement. The purpose of this paper is twofold: first to provide descriptions of the implementation components associated with each of the 12 principles identified in phase 1, capturing the potential variability of implementation strategies across different institutions and educational settings; second, to gain insight into enablers and barriers related to the implementation of programmatic assessment.

Methods

Participants and data collection

A sample of 15 programmes from Healthcare Professions Education that had implemented aspects of programmatic assessment was identified by two of the authors (CVD, AF). The sample included undergraduate and postgraduate programmes from six different countries across three continents. Following the Ottawa 2020 conference, the participants in the group who achieved consensus on the

12 principles of programmatic assessment (Heeneman et al. 2021), plus one additional programme, were invited to complete a 16-item questionnaire about implementation of programmatic assessment.

The survey was a semi-structured questionnaire distributed to participants as a Google form comprised of open-ended questions. Participants were asked to describe how they implemented each of the twelve principles of programmatic assessment in their own programmes, providing, where possible, specific examples. In addition, participants were asked about enablers, barriers encountered, lessons learned, and unintended consequences of the implementation. Information on the level of training (undergraduate vs graduate) and programme discipline (medicine, dentistry, veterinary) were also collected. The survey was developed by the authors and vetted by a subgroup of assessment specialists participating in the study.

Data analysis

Survey responses were analysed using a deductive thematic analysis (Boyatzis 1998; Braun and Clarke 2006). Given that the twelve principles of programmatic assessment share similar theoretical underpinnings (Heeneman et al. 2015; Torre et al. 2020) and have a high degree of interrelatedness, the programmatic assessment principles were used as a framework (or sensitizing concepts) for the thematic analysis (Brooks et al. 2015). Using an iterative collaborative thematic analysis process, two of the authors (DT, NR) read through the data looking for emergent new themes, met to discuss the themes developed, and then further refined them to build a common understanding of the data and identify relationships within and across themes. Ultimately, data were organised into three overall meaningful clusters (Figure 1).

Results

Implementation data were collected from 15 programmes (eight undergraduate (primary degree) medical education, three postgraduate (residency or vocational) medical education, one undergraduate dentistry and two undergraduate veterinary programmes) (Table 1). All but one of the programmes involved in the consensus statement survey, also took part in the implementation survey.

We present our findings in two parts. Firstly, we describe the implementation of programmatic assessment principles across different institutions reporting on implementation components that were common across programmes (mentioned by more than 50% of respondents) and provide examples of how the principles were translated into educational practices (Part 1). Secondly, we present findings on enablers, barriers, lessons learned, and unintended consequences of the implementation of programmatic assessment (Part 2).

For each theme, we present a table which identifies implementation components related to principles, and then provide specific examples from five programmes covering different countries and contexts. Our aim is to present a comprehensive and meaningful spectrum of implementation descriptions that captures the variability of implementation data in different contexts. It is important

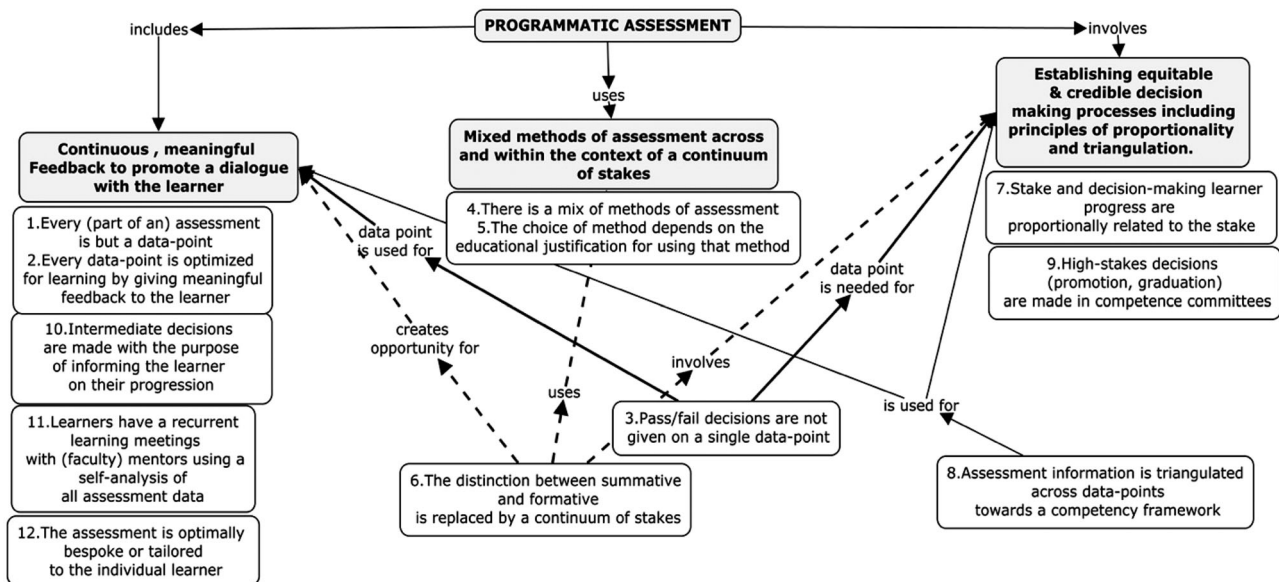


Figure 1. Principles of programmatic assessment grouped by three implementation themes.

Table 1. Programmes participating in part 2 PA consensus statement by discipline, level of training, programme size, and country.

Discipline	Undergraduate (UG) or post-graduate (PG)	Number of learners	Country
Dentistry	UG	377/year	United Kingdom
Medicine	PG	462	Australia
Medicine	PG	136	Canada
Medicine	PG	38	Canada
Medicine	UG	360/year	Australia
Medicine	UG	200/year	Australia
Medicine	UG	260/year	Canada
Medicine	UG	288/year	Canada
Medicine	UG	50/year	Netherlands
Medicine	UG	350/year	Netherlands
Medicine	UG	300/year	New-Zealand
Medicine	UG	280/year	United Kingdom
Medicine	UG	32/year	USA
Veterinary	UG	30/year	Canada
Veterinary	UG	225/year	Netherlands

to note, while we have included specific principles under each of our three themes, we acknowledge that all principles in the model are interconnected and function as interacting parts of a larger programmatic assessment system.

Part 1 – implementation of programmatic assessment principles

The implementation of the 12 principles of programmatic assessment were grouped under three themes: (1) Continuous and meaningful feedback to promote a dialogue with the learner for the purpose of growth and development; (2) Mixed methods of assessment across and within the context of a continuum of stakes; and (3) Establishing equitable and credible decision-making processes including principles of proportionality and triangulation.

Theme 1: Continuous and meaningful feedback to promote a dialogue with the learner for the purpose of growth and development:

There were nine implementation components (Table 2) aligned with principles 1, 2, 10, 11 and 12 under this theme.

Most programmes reported an implementation in which every assessment contributed a single data-point optimised for feedback, whether that was occurring in real time or at specific intervals. The delivery of narrative feedback using

aggregated data from multiple assessments was often mentioned. Feedback was delivered to learners in verbal and/or written form, with the intent of providing support for reflection and to guide future learning. The majority of programmes highlighted the importance of developing a relationship between supervisor and learner to enhance feedback quality and foster a process of meaningful reflection and goal setting.

All programmes reported the implementation of a system in which learners were assigned a coach or advisor to support their progress and achievement. The majority of programmes indicated that coaches had recurrent meetings with learners (quarterly or biannually); however, in some instances, meetings were more ad hoc in response to poor academic performance or concerns being flagged. Most programmes used some type of electronic portfolio to gather, store, and readily access assessment data to facilitate faculty-learner transactions. Intermediate reviews were performed for the purpose of assessing progression, promoting learners’ reflection about their longitudinal learning trajectory, formulating learning plans and and/or additional remediation if needed.

Most programmes reported learners’ engagement and participation in the assessment process upon entering the program. In some cases, programmes required learners to respond to feedback and develop individual learning plans, with an emphasis on self-reflection. In some programmes, teaching was tailored to meet individual learning priorities whilst encouraging learners to take agency for their own development. Generally, however, this principle was only partially implemented due to the logistical and organisational challenges of individualising learning for a high number of learners.

Theme 2: Mixed methods of assessment across and within the context of a continuum of stakes:

There were four common implementation components (Table 3) aligned with principles 4, 5 and 6 under this theme.

All programmes indicated that no single type of assessment can capture all competencies and that implementations involved the use of multiple assessment methods and tools.

Table 2. Continuous and meaningful feedback to promote a dialogue with the learner for the purpose of growth and development – implementation examples.

Principles	Implementation Components	UG dentistry, EUR	PG medicine, NA	UG medicine, NA	UG medicine, EUR	UG medicine, AUS
P1: Every (part of) an assessment is but a data-point	Assessment data-points are collected, stored and accessible to faculty and learners Assessment data-points are mapped to a blueprint/competency framework	Workplace-based assessment data is provided to students daily ... there are over 150 possible observations organised into a workflow (e.g. History gathering; Examination; etc) ... staff routinely cover most of the expected observations in every session.	The central component of our in-training assessment system is the provision of daily written feedback with the use of field notes in the clinical setting based on the skills described in our competency framework. It is a formative process not involved in high-stakes decisions.	Written assessments are frequent and provide feedback on progress on short periods of curriculum, while the standard for written assessments is based on aggregate of these within a whole 'course' (e.g. few months).	All assessment data in the ePortfolio are aggregated and feedback is presented according to an assessment blueprint (knowledge tests) and/or an overarching competency framework.	Data from different assessments are tagged in a way that allows cross-format comparison (with an overarching organising taxonomy).
P2: Every data point is optimised for learning by giving meaningful feedback to the learner	A portfolio or ePortfolio is used to gather feedback Areas for improvement are flagged up informing individual learning plans	Staff make observations against a standardised descriptor and provide written and verbal coaching Feedback that is aligned to developing independence of the learner and to support effective student reflection. Data is displayed to the learners using an ePortfolio system to indicate where they need to focus their efforts.	The completion of field notes requires observation, a reflective discussion between faculty and learner and then documentation for future use. Learners also regularly receive feedback on all other assessment data which is triangulated to inform a bi-annual review which includes a reflective exercise by the learner, resulting in a learning plan.	Even though standards are not set at individual data-points each point yields feedback. If, on a data-point, a learner is flagged as 'checked in' with by faculty to identify if there are academic or non-academic issues, and whether support in either domain is needed.	All feedback forms/assessment forms provide ample opportunity to write narrative feedback. Likert scale response anchors are formulated as rubrics (standardised written feedback), reflecting progress over time.	We use electronic data collection for our OSCEs, progress tests and Situational Judgement Tests. This allows us to provide feedback with commonality across multiple assessments (e.g. feedback on diagnosis skills might be provided across MCQ and OSCEs)
P10: Intermediate review is made to discuss and decide with the learner on their progression.	Faculty are trained to give meaningful feedback following rubrics/standard operating procedures Assessment data is reviewed regularly by faculty to trigger remediation support where required	Staff receive training in mentoring. The academic advisor system also supports students with 6-monthly feedback from a clinical progress panel, and the results for formative assessments (OSCE, SBA, SAQ etc.) are all timed to coincide with academic advisor meetings ...	Field notes are organised and summarised with the learner and a narrative in Training Assessment Report is generated as an intermediate step. Learners failing to progress as expected will be flagged and provided help to return to trajectory. Final summative decisions are made by a local competency committee.	A progress review document and individual learning plan is uploaded onto an e-portfolio and is signed off online by the coach. Coaches, who take part in this intermediate review, are not the ones that decide on progression (high-stakes decisions)	Following intermediate student/mentor meetings action plans for remediation are implemented where necessary. This is done in consultation with study advisors, counsellors, and/or others as necessary.	Prior to the end of the year, faculty consider learners who fail to achieve a satisfactory standard in their assessments, and these learners are given feedback, guidance and offered further learning opportunities.
P11: Learners have recurrent learning meetings with (faculty) mentors/coaches using a self-analysis of all assessment data.	A faculty mentoring/coaching/academic tutoring system is in place Mentors and learners need to meet regularly to positively engage staff and students in the process	These approaches are supported by a robust academic advisor system that requires students to analyse and reflect upon their data through the creation of action plans using SMART goals. Evidence of engagement with this process also forms an important part of the progression criteria ...	The most meaningful feedback depends on the ability of the learner/preceptor team to develop a reflective discussion that promotes learner development by stimulating the learner to accept guidance for their self-assessment. This is a central concept to our system.	We have twice yearly progress review sessions between students and coaches using an R2C2 model of facilitated feedback. Learners meet with their Coach who has reviewed the learning plan and ePortfolio and have a facilitated feedback dialogue.	Students meet with their mentor 3-4 times a year. During these mentor meetings, portfolio evidence is reviewed and discussed, followed by formulation of new learning plans.	Students have recurrent (weekly) meetings with clinical skills faculty coaches in years 1 & 2, but less so in years 3 & 4. Coaches remain consistent across the year and provide feedback on clinical skills development and help students plan future learning.
P12: PA seeks to gradually increase the learner's agency and accountability for their own learning through being tailored to support individual learning priorities	Students are encouraged or required to reflect on feedback (narrative/real-time/written) in self-directed learning	We have a longitudinal component to support the personal development of learners through establishing and enhancing tacit psychosocial skills such as self-efficacy and reflection. We use milestones (competencies) for learners to evidence personal development.	Learners are encouraged to engage in the process with reflection. The overriding goal is to teach our [learners] to become reflective learners seeking guidance for their self-assessment.	Students review electronic portfolio in advance [of tutor meetings] and reflect on achievement of competencies and personal learning plan.	Students have to take all assessment data into account when reflecting on performance and competence development and when formulating new learning plans.	This is much harder for us to implement in years 3 & 4 when our students are increasingly dispersed, and our course becomes more modularised.

UG: undergraduate; PG: post graduate; NA: North America; EUR: Europe; AUS: Australia.

Table 3. Mixed methods of assessment across and within the context of a continuum of stakes: implementation examples.

Principles	Implementation Components	UG medicine, EUR	UG medicine, NA	UG medicine, NA	UG medicine, EUR	UG medicine, AUS
P4: There is a mix of methods of assessment.	Multiple types of assessments employed; e.g. workplace-based assessments / knowledge tests (progress testing) / simulated assessments (OSCEs) / licensure examinations / written assessments / self & peer assessments	Knowledge is assessed using MCQ examinations (progress testing) with high reliability; clinical skills are assessed using a wider mix of assessment tools, including workplace-based assessments (miniCEXs, DOPs, longer case-based discussions etc) with high face validity, and using simulated observational assessments (e.g. OSCEs). Individual and group assessment also includes written work, oral and poster presentations.	Students complete several, low-stakes OSCEs in Years 1-4. Learners receive narrative feedback from many faculty members (longitudinal preceptors, PBL facilitators, communication skills preceptors, physical diagnosis preceptors, clinical teachers, research preceptors; journal club facilitators, research faculty, etc.) and also from peers (during research experiences and in PBL).	Mixed assessment methods are used: Written Exams (including progress tests; Workplace Based Assessments which include direct observation (linked to EPAs) in clinical rotations, mid unit formal feedback for core clinical rotations, and end of rotation / end of elective assessments; Reflective Portfolios which involve small group and narrative reflections; OSCEs assess the demonstrated competencies/ learning outcomes in clinical and communication skills.	Programme contains mixture of assessments including self and peer assessment. Our programme uses: -standardised knowledge tests (key feature, progress test) -written assignments on various topics -case presentations / oral presentations -research projects -workplace-based assessments (e.g. mini-CEX, DOPS, 360-degree feedback, fieldnotes)	We use a mix of MCQs, SAQs, essays, reflective pieces, professionalism standards, attendance requirements, SJTs, OSCEs, Long Case, Standardised Case Based Discussion, MiniCEX, 360 degree feedback.
P5: The method chosen should depend on the educational justification for using that method.	The assessment mix is designed to align with and cover the competence-map / assessment blueprint	Assessment tools are fitted to the type of learning being assessed. For example, progress testing for longitudinal tracking of applied knowledge growth. More nuanced continuous assessment includes peer-assessed in-class work, maintaining clinical logs in a portfolio, and self-reflection in group and tutor meetings.	Students complete several, low-stakes OSCEs in Years 1-4. Additionally, students complete weekly MCQs to self-assess knowledge and weekly essays to demonstrate application and integration of knowledge. All of our assessments map onto one or more of our nine competencies.	The choice of assessments in the mix are based on published guidelines available to both faculty and students. Recommendations are made by Regional Committees and final decisions are made by the Promotions Board. Each modality must be passed for progress or promotion after sufficient academic support has been offered.	Learning drives the assessment. When designing the programme, we started by identifying learning tasks to support competence development. Only then did we engage in design of appropriate feedback modalities / assessment formats – to optimise constructive alignment.	We still run our programme within a fairly traditional University assessment culture and so need to adhere to the requirements there, but the Uni as a whole is moving towards a reform of assessments and an increased focus on learning so this is helping us (especially in recent years).
P6: The distinction between summative and formative is replaced by a continuum of stakes.	Active narrative change-management to move away from formative v summative and pass v fail to continuum of stakes and readiness to progress Alternating multiple high and low-stakes assessments while informing students about the stakes of the assessments	Our emphasis is on continuous, low-stakes assessments, with frequent feedback and rapid remediation. We are replacing the narrative of pass/fail with that of 'requiring self-directed learning in the area of xxx before your next assessment, to try and shift learners' attitudes to a continuum way of thinking, away from discrete assessment decisions – this has shown some success so far. We are moving to using sequential lower-stakes termed mini-OSCE examinations in the assessment of clinical skills to replace high-stakes end of year OSCE examinations.	We developed the nomenclature of formative and summative portfolios when implementing our assessment system. We do not have any grades or class ranks throughout the five-year programme. High-stake decisions occur at the end of stages. All other assessments are low-stakes, beyond those required by regulatory authorities (USMLE) Many undergraduate medical programmes would struggle with principle 6 due to the reliance on high-stake tests and assignment of grades or class ranks. Programmatic assessment requires an assessment for learning philosophy and appropriate resources.	A continuum of stakes is evident within assessment modalities and ranges from those geared predominantly for learning and feedback, such as direct observations, to those which have a higher stake for progress and promotion decisions such as end of rotation assessments Assessment information is triangulated within assessment modality frameworks. Failure in one modality cannot be compensated by doing well in another, as all competencies must be achieved and assessed at the appropriate level possible within the Miller's Pyramid.	Given that high-stakes decision-making is to be based on multiple and aggregated data, students are informed about the stakes of single assessments. Central governance as well as involvement of all key stakeholders in design and implementation is essential. Tensions may arise, however, if programme leaders or boards of examiners feel pressured by accreditation bodies (for example), or in case of legal / regulatory constraints. Lots of attention needs to be paid to personal assessment beliefs; learners and faculty often see low-stakes as high-stakes.	We have multiple assessments with graduated stakes, spread throughout our entire course. There are higher stakes assessments towards the end of each semester, but many prior assessments (in the same or similar format) with smaller mark allocations, leading up to the point of decisions about progress to the next year. Progress to the next rotation is considered a lesser decision when compared to progress to the next year level, which is a lower stake than graduation from the course, so the decision-making and summative nature of the assessments would be proportionally considered.

The variety of assessments implemented allowed for the gathering of both numerical and narrative data that could be used to provide rich and meaningful feedback to learners. Nearly all programmes mentioned that the mix of assessments was designed against a competency framework, and many indicated adherence to constructive alignment and a programme of assessments designed using Miller's pyramid. Most programmes indicated multiple low-stakes workplace-based assessments as particularly important for assessing real-world situational competence, and most reported the use of simulated assessments (e.g. OSCEs) and written assessments (including progress testing) in the knowledge domain.

Most programmes reported the implementation of a continuum of stakes by the use of multiple low-stakes assessments with frequent delivery of feedback (e.g. direct observation workplace-based assessments or sequential mini-OSCEs) interwoven with high-stakes assessments performed at key developmental transition points in the curriculum (e.g. progression exams before entering clinical rotations, final written exams and OSCEs or national licensing exams). This principle had a wide range of implementations, related to timing and progressive combination of low- and high-stakes assessment. Furthermore, providing students with information about the stakes of assessment and explaining how such a continuum of stakes was going to be operationalised was a frequent implementation strategy in the majority of programmes.

Many programmes highlighted the importance of providing information and guidance to encourage stakeholders to think of assessments as being on a continuum of stakes rather than pass/fail. The implementation of multiple low-stakes assessments with feedback prior to high-stakes assessments allowed programmes to begin shifting from a predominant summative assessment approach to a language and culture of assessment for learning through a continuum of stakes.

Theme 3: Establishing equitable and credible decision-making processes including principles of proportionality and triangulation:

There were ten common implementation components (Table 4) aligned with principles 3, 7, 8 and 9 under this theme.

For most programmes, high-stakes decisions (e.g. decisions to progress to the next year of training) were made by a committee based on the review and aggregation of multiple assessment data points collected over time. The timing of high-stakes decisions varied by programme, with some occurring yearly and others biannually or quarterly. Low-stakes decisions (e.g. a decision on a performance of a direct observation of clinical skills) were based on a single or few assessment data points for the purpose of providing feedback to promote learners' growth and development. However, the majority of programmes indicated that when implementing the principle of proportionality, low-stakes assessments would be one of the data points that contributed to a high-stakes decision.

Most programmes reported that implementation of high-stakes decisions involved some form of a committee and described some version of a two (or more) stage committee decision-making process (e.g. a progress committee makes a first determination about a learner's performance after

reviewing all assessment data, and then an examiner committee reaches a final decision). The typically described committee consisted of a group of independent, trained examiners who review learners' assessment data, often collected in an e-portfolio, and then engage in a data-driven deliberation to reach a final decision. Most programmes highlighted the need for well communicated, pre-determined procedures, rules or rubrics to guide decision-making and to ensure transparency and credibility, often with direct learners' engagement and participation in a shared decision-making process.

Triangulation was reported by most programmes as important in the high-stakes decision-making process and was generally implemented by synthesizing data from multiple assessment modalities that had been previously mapped to an overarching competency-based framework.

Whilst all programmes described decision-making processes that included the monitoring of longitudinal performance of all learners, the majority reported that most decisions were clear-cut and did not require much, if any, discussion, with outcomes not coming as 'surprises' for faculty or learners. The majority of programmes reported that in depth committee discussions usually focused on a relatively small number of learners for whom there were concerns over meeting performance standards.

Part 2 -enablers, barriers, and lessons learned

There was agreement that the strongest enabler for programmatic assessment implementation is strong leadership with a clear vision supported by committed faculty. The majority of programmes mentioned that having clearly communicated, accessible goals and delivery plans was an essential enabler. Several programmes mentioned that investing in the development of a design and implementation plan was fundamental to the success of the implementation.

A centralised oversight coupled with the creation of an assessment working group who would promote the philosophy of programmatic assessment and engage and communicate with all stakeholders were key enablers in the implementation process. Having a clear vision of the assessment system, involving learners in the assessment process, establishing assessment goals, and implementing defensible processes and procedures were all mentioned as necessary steps to enable successful implementation.

Another common enabler (and barrier) was the level of buy-in and engagement of all stakeholders in the design, development and delivery of the implementation process. Fundamental to the success was an ongoing commitment of resources to train faculty and learners on the principles, operationalisation of programmatic assessment, coupled in the longer term, with a commitment to invest in software (e-Portfolios) to support feedback delivery, learners' agency, and triangulation procedures for decision-making. The creation of an ongoing faculty development programme as part of the implementation, focusing on assessment literacy, a good knowledge of the competency framework in use and providing feedback for learning, was a critical enabler. Such effort clearly involved the leveraging of resources and leadership support. Investing resource into developing a strong coaching system to evaluate and support learners' progress and achievement was also mentioned.

Table 4. Establishing equitable and credible decision-making processes including principles of proportionality and triangulation – implementation examples.

Principles	Implementation Components	UG vet science, EUR	UG medicine, EUR	UG medicine, NA	PG medicine, NA	UG medicine, AUS
P3: Pass/fail decisions are not given on a single data-point.	High-stakes decisions are not made on single assessments / tasks High-stakes decisions are usually made once at the end of a stage of learning	No high-stakes/summative judgments are made based on a single assessment data-point.	High-stakes decisions are made once a year, based on aggregated data, so all data is used.	Longitudinal themes all get considered and inform decision-making at the end of a course. Stakes of written assessments are aggregated, not individual.	No pass/fail decision is made on one data point. A single bad score could FLAG a learner for follow-up. We look for patterns and defer decision-making until we have enough data.	Almost never is a single data point the basis for a pass/fail decision for the year. For every component there are repeated assessments which are combined to make a decision.
P7: Decision-making on learner progress is proportionally related to the stakes.	Single Lower-stakes assessments are used for feedback yet they are all ultimately aggregated and used to inform high-stakes decisions There are 'external' required high-stakes assessments (licensure, qualification exams) to consider in the decision	Learners collect feedback using multiple low-stakes WBAs which are aggregated at the competency level using an e-Portfolio. Intermediate mentor reports based on frequent individual mentee-mentor meetings form an important information source for the assessors in the committee making high-stakes decisions.	Low-stakes assessments are optimised for learning, with no direct pass/fail consequences but all do have a 'stake', as these are aggregated for the high-stakes decision.	Within the confines of university regulations, high-stakes decisions are not made explicitly on 'competencies', but on all data points, by modality mapped to competencies, within a course and year. Over time we have found that lower-stakes flags are very predictive of learners who will be in difficulty on high-stakes review.	Assessments are low-stakes until we use the longitudinal data to decide if we recommend progress to exams, or to hold a trainee back - where a case must be made from the data, across multiple domains OR based on repeated problems within a single problematic area.	For most students, who have sufficiently passed all exam modalities, no deliberations are held in exam boards. For students about whom doubts are flagged in low-stakes assessments, extensive deliberations can be held.
P8: Assessment information is triangulated across data-points towards an appropriate framework.	Careful triangulation of data to align with competence-map is important for equitable and credible decision-making Standard operating procedures / rubrics / pre-defined rules (e.g. on aggregation of data) is required for transparency Assessment data needs to be holistic and unambiguous, commonly by using a portfolio	Two assessors independently judge learner performance by triangulation of quantitative and qualitative assessment data in a digital portfolio following a defined rubric. The assessors reports on strengths and weaknesses and a pass/borderline/fail decision. If there is disagreement between assessors or when the decision is borderline, the learner's portfolio is discussed by the committee.	Multiple structured assessment forms are generated following defined rubrics which give ample room for narrative feedback. This is collected in an e-portfolio, which supports aggregation of data and triangulation of data aligned to the competency framework outcomes for decision purposes, by generating overviews (partly customised, e.g. show all narrative feedback for the competency)	Data in an e-portfolio is reviewed against pre-determined standards made clear to learners. All context provided by faculty alongside longitudinal data visualization and history of focused learning plans is considered to inform decisions. We have developed flexible recommendation making in the face of decisions, within the confines of university regulations and using defined rubrics etc. (e.g. partial progress with conditions, etc.)	We examine data from multiple teaching groups (three hospitals with largely non-overlapping teachers), and from multiple rotations over time. Data points from all parts of the competency framework portfolio are considered in decision-making.	Triangulation happens when students are flagged and not passed. Deliberation incorporates review of the all assessment results and entries in a portfolio Careful decisions on progression or otherwise are made within the confinements of University policies. There is a need for transparent procedures.
P9: High-stakes decisions (promotion, graduation) are made in competence committees.	Multiple stage decision-making processes are often employed ... e.g. small group meetings followed by final competence committees followed by exam-boards Decision-making 'committees' need be trained, experienced and diverse Discussions usually focus on individuals - no 'surprises' in decision-making	High-stakes decisions are made by two assessors (members of an extensively trained committee of assessors). All members from the portfolio examination committee are senior employees at and are trained in assessing e-portfolios.	A mentor gives intermediate advice on progress/performance half-way through and at the end of the year. All information (portfolio and mentor advice) is discussed in a plenary meeting of mentors which generates another advice. All this data is then reviewed by a portfolio assessment committee who make the high-stakes decision] This procedure is described in guidelines and explained in a lecture, and to by the mentor to the learner.	Student Progress Committees (SPC) are composed of people who are not in key course director or support roles and who have extensive relevant experience. A final higher body Board of Examiners makes final decisions (e.g. regarding the consequences of failure) and hears appeals. Almost always they uphold the decision of SPC, unless new information is brought to bear.	We have a competency committee (CC) trained in identifying trends. For individual cases arguments are presented and case data is re-examined, and a recommendation is made to a separate training committee and the programme director. This occurs toward the end of each year. The decisions of the CC should not be a surprise as the learner receives various forms of assessment/feedback throughout.	Decision-making committees can be a way in which fairness, equity and credibility are achieved, but not the only one. A committee of people need to understand the principles of programmatic assessment. Committees of experts need information to not be lacking in content or agreement between assessors in order to make informed defensible decisions.

Where programmes had implemented programmatic assessment in new programmes or in conjunction with a full curriculum review, it was evidently a smoother process than in programmes which were attempting to integrate within established systems. Careful constructive alignment with detailed curriculum mapping was noted as being easier in new implementations.

The commonly mentioned barriers to implementation to a large extent mirrored the enablers. The level of organisational commitment to change, demands on faculty time, and availability of resources were cited by all responders as barriers, but not necessarily insurmountable ones. The biggest barrier to successful implementation was a general resistance to change, with most programmes mentioning the difficulty in changing embedded educational and assessment cultures, expectations and attitudes – of faculty, educators and learners. A shift in the mindset and culture of assessment among learners was identified as a significant barrier, as was a lack of planning and provision of resources for faculty professional growth. The lack of strong leadership committed to making changes at the system level was recognised as a significant obstacle to implementation by many institutions.

Institutional policies and requirements from accreditation and regulatory bodies were reported as potential barriers. For example, the requirement on learners to pass a high-stakes national licensing exam at some point in their training was seen as counter-intuitive to the theoretical principles and as a hindering factor to the acceptability of and learners' buy-in to programmatic assessment. However, most programmes were able to make changes and implement programmatic assessment within the boundaries and regulations set by university policies and national accreditation bodies.

The implementation of programmatic assessment afforded programmes the opportunity to learn a number of lessons related to learners, faculty, and their educational system. Programmes learnt that involving learners in programmatic assessment innovation is just as important as buy-in from faculty members. Most programmes described that implementing programmatic assessment helped leverage the learning function of assessment and provided more opportunity to identify struggling learners early on, thereby allowing more time for improvement and for targeted and robust interventions. Furthermore, learners were more likely to seek feedback proactively and take ownership of their own learning while faculty were more comfortable with assessing, discussing, and making decisions about learners' progress and performance.

Many programmes described having learnt the importance of engaging regularly with stakeholders to maintain a shared vision, frequently communicating the why, what, and how change is going to occur. Most respondents reported that it took perseverance, strong leadership, and institutional commitment to implement and recognise the educational gains of programmatic assessment. Several respondents also recognised that understanding the contextual factors and the cultural landscape of the institution was vital to enact a process of change. Changes occurring in one particular context may not be feasible or as successful in another. Many programmes indicated that programmatic assessment is a fundamental disruption from traditional

assessment and requires a major paradigm shift; therefore, creating a context supportive of change is imperative for success.

Respondents indicated several unintended consequences of the implementation of programmatic assessment. For some programmes, implementing programmatic assessment provided a new insight into the curriculum and allowed identification of areas for improvement. The change to programmatic assessment enhanced feedback literacy and helped students understand how to best utilise the feedback they received from faculty.

One programme reported that instead of having multiple data points and few decision points in accordance with programmatic assessment and assessment for learning principles, for some reason, there was an increase in the number of decision points, resulting in students' perception of being part of an environment of constant summative assessment.

Some programmes reported an increase in faculty time and workload, particularly at the inception of programmatic assessment implementation, but noted that the longer-term benefits in terms of the better communication and dialogue between faculty (coaches) and learners outweighed the initial investment of time resource.

Several programmes indicated that their implementation of programmatic assessment coincided with an unexpected increase in the use of technology enhanced assessment; the use of an e-portfolio and the addition of portable technology to collect experiences allowed learners to share their work, promoting reflection and creating a safe environment to interact with faculty. Additional details on enablers and barriers are reported in [Supplementary Appendix 1](#).

Discussion

There were a wide range of implementation descriptions across different programmes, yet overall components of implementation, enablers and barriers were similar across disciplines and undergraduate and postgraduate programmes. The actual implementation of the 12 programmatic assessment principles remained, for the most part, faithful to the original enunciation and rationale. For example, most programmes had some form of assessment committee for high-stakes decisions; however, what the committee was named, what procedure was used to analyse data, or how often the committee met was found to have a range of implementations across programmes. The use of review committees for high-stakes decisions, the utilisation of mixed methods of assessments, the implementation of every assessment as a data point optimised for feedback, and the development of a continuum of stakes were just some of the principles translated into practice within a range of implementations across different programmes and educational contexts.

The critical role of implementation in programmatic assessment has been discussed (Bok et al. 2013), and it has been suggested that the quality of implementation defines the success of programmatic assessment (Van Der Vleuten et al. 2019). Key implementation elements include design based on principles and participation (Jamieson et al. 2017; Schuwirth et al. 2017), the role of a portfolio to integrate

assessments (Dannefer and Henson 2007), the role of a committee in high-stakes decision-making (Van Der Vleuten et al. 2015), students' perception of the assessment (Heeneman et al. 2015; Schut et al. 2018), the relationship between learners and faculty (Schut et al. 2021), and the parallel implementation of curriculum and programmatic assessment (Freeman and Ricketts 2010; Ricketts and Bligh 2011). More recently, Jamieson et al. (2021) found that a successful implementation of programmatic assessment transformed the supervisors' role and their views of assessment practices, whilst enhancing the supervisor-learner relationship. Adequate implementation of programmatic assessment is also critical to provide data to inform outcomes of competency based medical education (Hauer et al. 2018; Iobst and Holmboe 2020; Misra et al. 2021).

The implementation descriptions of the consensus group contribute to enhance our knowledge and understanding of the spectrum of implementation related to the principles of programmatic assessment. Implementation is a process not an isolated event (Hall and Hord 2015). Most programmes reported that the implementation process of programmatic assessment takes time (years not months), needs a strategic well-designed and well budgeted plan, requires strong leadership and support, and demands a shared vision of change and a culture shift across the entire educational enterprise. Further, it involves an ongoing programme of faculty development and coaching with continuous monitoring of progress and assistance to face and resolve challenges. Implementation of programmatic assessment requires sustained commitment of resources and ongoing vigilance to maintain the integrity of its key components.

Designing a plan, securing leadership support, allowing opportunities for flexibility and adjustments, a shared belief of relevance and value among stakeholders are all important factors reported in the managing change literature (Gale and Grant 1997; Schneider 2014). The descriptions of implementations of programmatic assessment are aligned with several of these factors, yet adaptation to context and its complexities seems to emerge as an essential ingredient for a successful implementation of the principles of programmatic assessment across programmes.

Assessment as a continuum of stakes (principle 6) had a particularly wide range of implementations. Many programmes implemented and executed plans to explain and educate learners about the meaning of this principle, including how it was going to be operationalised, promoting a culture of growth and improvement that, in some cases, led to the elimination of grades and class ranks. This approach created opportunities to foster learners' agency with the assessment process. This was aligned with previous evidence suggesting that a continuum of stakes is perceived by learners as complex and is strongly related to their perception of being able to control and take agency of the assessment process (Bok et al. 2013; Schut et al. 2018).

Learners' agency is a complicated process; it takes effort and requires mentorship and support to counteract the pressure of established social norms and cultural beliefs (Watling et al. 2021). Teacher-learner relationships play a critical role in learners' assessment perceptions; teachers who show a less dominant approach in the assessment

relationship enable learners' agency, promoting a positive learner perception of assessment for learning (Schut et al. 2020a). If teachers exert control over the low-stakes assessment process it can hinder the learning function of assessment and generate tensions in the teacher-learner relationship (Schut et al. 2020b). Such tensions can be relieved in the use of progress committees or clinical competency committees, where important assessment decisions have a collective responsibility rather than that of a single individual. The use of a progress or competency committee was a key component in almost all descriptions of the implementations of programmatic assessment.

In the implementation of programmatic assessment, fostering a feedback seeking behaviour with a learning goal orientation (focused on growth and improvement) rather than a performance goal orientation (focused on gaining a positive judgement and garnering recognition of their own ability or on avoiding negative feedback) (Bok et al. 2013; Teunissen and Bok 2013) may promote learners' acceptance and affect their perceptions of assessment as a continuum of stakes, ultimately facilitating the implementation of this principle. It has also been suggested that reflection is the path to acceptance of feedback that is out of line with learners' self-assessment (Sargeant et al. 2008).

One of the principles, (principle 12, assessment tailored to the individual learner), proved to be challenging to fully implement, particularly in programmes with a large number of learners. However, the implementation of the other principles did not seem to be significantly affected by the size of a programme, at least to the extent that a principle could not be at all implemented because of a large programme size.

One of the widely recognised barriers to implementation was the need for a paradigm shift in assessment culture within the organisation, amongst leaders, learners and educators. Programmes which have longer-standing established implementations reported more success whereas newer implementations are still experiencing this cultural shift, and there were noticeable differences in the extent of this across contexts. Understanding the organisational and cultural context of programmes is essential for the implementation and sustainability of any intervention or change because all stakeholders are an integral part of a shared environment (Damschroder et al. 2009; Marks et al. 2010). Context consists of a number of interacting variables and set of circumstances that surround a specific implementation. All parts which shape the context of a system need to be legitimised and involved.

The emerging concepts of self-theories which impact personal motivation and growth mindset, built on the social cognitive model proposed by Dweck (Dweck and Leggett 1988) can be considered in relation to stakeholders' attitudes to changing culture. The way in which self-theories may impact on Health Professions Education has been explored (Teunissen and Bok 2013), and the model has been generalised to organisations (Canning et al. 2020). External attributes (such contextual education factors) may be considered as 'fixed' (entity theory) or 'adaptable' (incremental theory) depending on the mindset of the stakeholder. If stakeholders are incremental theorists, they are more likely to positively embrace change by seeking improvement and believing that external attributes can

change for the better. Conversely, if stakeholders are entity theorists, they will be more resistant to engage with or initiate change.

Programmes that endorse and adopt a growth mindset culture will motivate learners and educators to take on more challenges and initiate change, encourage stakeholders to view 'failures' as learning opportunities and inspire confidence in innovation (Dweck 2019; Canning et al. 2020). Fostering an incremental theory approach and a learning goal orientation within individual learners and teachers, and at the organisational level would be beneficial for the implementation of programmatic assessment. There are some encouraging results in the literature about potential educational interventions to promote a growth mindset (Aronson et al. 2002; Blackwell et al. 2007). However, more research is needed to understand how to promote a growth mindset more effectively and systematically among individuals and across organisations.

Conclusions

The majority of the theoretical programmatic assessment principles achieved by consensus were successfully implemented by most programmes. A few principles like using a true continuum of stakes (principle 6), proportionality about decision-making processes (principle 7) or procedures for triangulation of data (principle 8) had a wider spectrum of implementations. Principle 12 (assessment tailored to the individual learner) proved to be the most challenging, but not impossible, to fully implement. Overall a description of the implementation of programmatic assessment across multiple educational contexts provided new insights to gain a better understanding of programmatic assessment and shed some light on how it can be best operationalised to achieve its full educational benefits. Implementation of programmatic assessment takes time, needs strong leadership commitment and support, and involves a continuous effort at the individual and organisational level. A paradigm shift from the mindset of a traditional assessment approach to that of a programmatic assessment culture by faculty, and students, coupled with the constraints posed by university policies and procedures, were significant yet not insurmountable barriers to implementation. Programmatic assessment implementation provided a new insight into the curriculum, allowing programmes to identify areas for improvement. It also enhanced feedback literacy among students and faculty, improved feedback quality, and allowed early identification and support of struggling learners. Successful implementations of the principles of programmatic assessment (mix of methods of assessment, triangulation, longitudinality, and proportionality) improve the data available for decision making processes, whilst keeping the amount of data and documentation manageable and meaningful.

Recommendations for future work

There are a number of recommendations for future research. First, investigate further the critical factors that may influence the implementation of specific principles in order to enhance their implementation. The role of required high-stakes assessments conducted by external regulatory bodies

(licensure, qualification exams) on the implementation of specific principles needs to be further determined and explored. Second, inquire about the development of effective faculty development programmes that may enhance the quality of implementation, particularly about key aspects of programmatic assessment such as feedback delivery and programmatic assessment literacy. Further, ways to promote students' understanding, adoption, and utilisation of programmatic assessment practices should be sought. This could include examination of explicit and hidden curriculum messages that might cause learners cognitive dissonance, for example the reality of high-stakes consequences of professional training programmes, such as success in residency matches. Third, additional inquiries should be conducted about key contextual and cultural factors that affect implementation and how such factors influence implementation across different institutions worldwide. Finally, more research is needed to evaluate the impact of a complete and effective implementation of programmatic assessment, learner competency and wellness, and the extent to which this approach ultimately impacts health care outcomes at the patient and system level.

Acknowledgements

This work was made possible thanks to the efforts of individuals, including the authors, who took the time to complete the survey questionnaire describing implementation experiences in their programmes. As such, the authors would like to acknowledge and thank the following: Marjan Govaerts, (Maastricht, the Netherlands) Lambert Schuwirth (Flinders University, Australia), Teresa Chan (McMaster, Canada), Paul Dilena (General Practice, Parkside, Australia), Vincent Arockiasamy (University of British Columbia, Canada) and Kent Hecker (University of Calgary, Canada).

Disclosure statement

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

Notes on contributors

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.

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

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

Harold Bok, Faculty for Veterinary Medicine, Centre for Quality Improvement in Veterinary education, Department 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.

Beth Beirer, MD, Director of Assessment and Evaluation, Cleveland Clinic Learner College of Medicine of Case Western University, Cleveland, OH, USA.

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.

Tom Laughlin, MD, Associate Professor of Family Medicine, Department of Family Medicine, Dalhousie University, Halifax, Canada.

Kieran Veerapen, MD, Associate Professor of medicine, Assistant Dean for Faculty Development, University of British Columbia, Vancouver, BC, Canada.

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.

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.

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

ORCID

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

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

Beth Bierer  <http://orcid.org/0000-0002-7952-8822>

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

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

References

- Aronson J, Fried CB, Good C. 2002. Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *J Exp Soc Psychol.* 38(2):113–125.
- Blackwell LS, Trzesniewski KH, Dweck CS. 2007. Implicit theories of intelligence predict achievement across an adolescent transition: a longitudinal study and an intervention. *Child Dev.* 78(1):246–263.
- Bok HG, Teunissen PW, Favier RP, Rietbroek NJ, Theyse LFH, Brommer H, Haarhuis JCM, van Beukelen P, van der Vleuten CPM, Jaarsma DADC. 2013. Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Med Educ.* 13:123.
- Bok HG, Teunissen PW, Spruijt A, Fokkema JP, van Beukelen P, Jaarsma DA, van der Vleuten CP. 2013. Clarifying students' feedback-seeking behaviour in clinical clerkships. *Med Educ.* 47(3):282–291.
- Boyatzis RE. 1998. *Transforming qualitative information: thematic analysis and code development.* Thousand Oaks (CA): Sage.
- Braun V, Clarke V. 2006. Using thematic analysis in psychology. *Qual Res Psychol.* 3(2):77–101.
- Bierer SB, Dannefer EF, Tetzlaff JE. 2015. Time to loosen the apron strings: Cohort-based evaluation of a student-driven remediation model at one medical school. *J Gen Intern Med.* 30(9):1339–1343.
- Brooks J, McCluskey S, Turley E, King N. 2015. The utility of template analysis in qualitative psychology research. *Qual Res Psychol.* 12(2): 202–222.
- Canning EA, Murphy MC, Emerson KT, Chatman JA, Dweck CS, Kray LJ. 2020. Cultures of genius at work: organizational mindsets predict cultural norms, trust, and commitment. *Pers Soc Psychol Bull.* 46(4): 626–642.
- Cook DA, Brydges R, Ginsburg S, Hatala R. 2015. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Med Educ.* 49(6):560–575.
- Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. 2009. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement Sci.* 4(1):50–55.
- 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.
- De Jong LH, Bok HG, Kremer WD, Van Der Vleuten CP. 2019. Programmatic assessment: Can we provide evidence for saturation of information? *Med Teach.* 41(6):678–682.
- 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(3):226–231.
- Dweck CS, Leggett EL. 1988. A social-cognitive approach to motivation and personality. *Psychol Rev.* 95(2):256–273.
- Dweck CS. 2019. The choice to make a difference. *Perspect Psychol Sci.* 14(1):21–25.
- Freeman AC, Ricketts C. 2010. Choosing and designing knowledge assessments: Experience at a new medical school. *Med Teach.* 32(7): 578–581.
- Gale R, Grant J. 1997. AMEE Medical Education Guide No. 10. Managing change in a medical context: guidelines for action. *Med Teach.* 19(4):239–249.
- Hall G, Hord S. 2015. *Implementing change: patterns, principles and potholes.* 5th ed. Upper Saddle River (NJ): Pearson.
- Hauer KE, O'Sullivan PS, Fitzhenry K, Boscardin C. 2018. Translating theory into practice: implementing a program of assessment. *Acad Med.* 93(3):444–450.
- Heeneman S, Oudkerk Pool A, Schuwirth LWT, Van der Vleuten CPM, Driessen EW. 2015. The impact of programmatic assessment on student learning: theory versus practice. *Med Educ.* 49(5): 487–498.
- Heeneman S, De Jong L, Dawson L, Wilkinson T., Ryan A, Tait G, Rice N, Torre D, Freeman A, van der Vleuten C. 2021. Ottawa 2020 consensus statement for programmatic assessment – 1. Agreement on the principles. *Med Teach.* DOI:10.1080/0142159X.2021.1957088.
- Jamieson J, Jenkins G, Beatty S, Palermo C. 2017. Designing programmes of assessment: a participatory approach. *Med Teach.* 39(11):1182–1188.
- Jamieson J, Hay M, Gibson S, Palermo C. 2021. Implementing programmatic assessment transforms supervisor attitudes: an explanatory sequential mixed methods study. *Med Teach.* 43(6):709–717.
- Iobst WF, Holmboe ES. 2020. Programmatic assessment: the secret sauce of effective CBME implementation. *J Grad Med Educ.* 12(4):518–521.
- Marks R, Clarke AM, O'Sullivan M, Barry MM. 2010. Context matters in programme implementation. *Health Educ.* 110(4):273–293.
- Misra S, Iobst WF, Hauer KE, Holmboe ES. 2021. The importance of competency-based programmatic assessment in graduate medical education. *J Grad Med Educ.* 13(2 Suppl):113–119.
- Norcini J, Brownell Anderson M, Bollela V, Burch V, Costa MJ, Duvivier R, Hays R, Mackay MFP, Roberts T, Swanson D. 2018. 2018 Consensus framework for good assessment. *Med Teach.* 40(11):1102–1109.
- Norcini J, Burch V. 2007. Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Med Teach.* 29(9):855–871.
- Ricketts C, Bligh J. 2011. Developing a "frequent look and rapid remediation" assessment system for a new medical school. *Acad Med.* 86(1):67–71.
- Sargeant J, Mann K, Van der Vleuten C, Metsemakers J. 2008. "Directed" self-assessment: practice and feedback within a social context. *J Contin Educ Health Prof.* 28(1):47–54.
- 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, 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.
- Schut S, van Tartwijk J, Driessen E, van der Vleuten C, Heeneman S. 2020a. Understanding the influence of teacher-learner relationships on learners' assessment perception. *Adv Health Sci Educ Theory Pract.* 25(2):441–456.
- Schut S, Heeneman S, Bierer B, Driessen E, Van Tartwijk J, Van Der Vleuten C. 2020b. Between trust and control: teachers' assessment conceptualisations within programmatic assessment. *Med Educ.* 54(6):528–537.

- 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(5):Doc56.
- Schuwirth LWT, Van der Vleuten CPM. 2011. Programmatic assessment: from assessment of learning to assessment for learning. *Med Teach.* 33(6):478–485.
- Schuwirth LW, van der Vleuten CP. 2012. Programmatic assessment and Kane's validity perspective. *Med Educ.* 46(1):38–48.
- Schneider J. 2014. Closing the gap ... between the university and schoolhouse. *Phi Delta Kappan.* 96(1):30–35.
- Sherbino J, Bandiera G, Doyle K, Frank JR, Holroyd BR, Jones G, Norum J, Snider C, Magee K. 2020. The competency-based medical education evolution of Canadian emergency medicine specialist training. *CJEM.* 22(1):95–102.
- Teunissen PW, Bok HG. 2013. Believing is seeing: how people's beliefs influence goals, emotions and behaviour. *Med Educ.* 47(11):1064–1072.
- Torre DM, Schuwirth LWT, Van der Vleuten CPM. 2020. Theoretical considerations on programmatic assessment. *Med Teach.* 42(2):213–220.
- Van Der Vleuten C, Schuwirth L. 2005. Assessing professional competence: from methods to programmes. *Med Educ.* 39(3):309–317.
- 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(3):205–214.
- 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 (NY): Routledge.
- Van Der Vleuten CPM, Schuwirth LWT, Driessen EW, Govaerts MJB, Heeneman S. 2015. Twelve Tips for programmatic assessment. *Med Teach.* 37(7):641–646.
- Watling C, Ginsburg S, LaDonna K, Lingard L, Field E. 2021. Going against the grain: an exploration of agency in medical learning. *Med Educ.* DOI:10.1111/medu.14532.
- Wilson M, Sloane K. 2000. From principles to practice: an embedded assessment system. *Appl Meas Educ.* 13(2):181–208.