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


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

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 postgraduate 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
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/or additional remediation if needed.

Part 2 – 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 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.

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

There were seven common implementation components (Table 4) aligned with principles 7, 8, 9 and 13 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.
### Table 2: Continuous and meaningful feedback to promote a dialogue with the learner for the purpose of growth and development – implementation examples.

<table>
<thead>
<tr>
<th>Principles</th>
<th>Implementation Components</th>
<th>Components</th>
<th>UG dentistry, EUR</th>
<th>PG medicine, NA</th>
<th>UG medicine, NA</th>
<th>UG medicine, EUR</th>
<th>UG medicine, AUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1: Every part of an assessment is but a data-point</td>
<td>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.</td>
<td>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.</td>
<td>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).</td>
<td>All assessment data in the ePortfolio is aggregated and feedback is presented according to an assessment blueprint (knowledge tests) and/or an overarching competency framework.</td>
<td>Data from different assessments are tagged in a way that allows cross-format comparison (with an overarching organising taxonomy).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P2: Every data point is optimised for learning by giving meaningful feedback to the learner</td>
<td>A portfolio or ePortfolio is used to gather feedback Areas for improvement are flagged up informing individual learning plans</td>
<td>Staff make observations against a standardised descriptor and provide written and verbal coaching feedback that is aligned to the development 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.</td>
<td>Even though standards are not set at individual data-points each point yields feedback. If, on a data-point a learner is flagged as possibly in difficulty, they are ‘checked in’ with by faculty to identify if there are academic or non-academic issues, and whether support in either domain is needed.</td>
<td>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.</td>
<td>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 OSCE)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10: Intermediate review is made to discuss and decide with the learner on their progression.</td>
<td>Faculty are trained to give meaningful feedback following rubrics/standard operating procedures</td>
<td>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, SNO etc.) are all timed to coincide with academic advisor meetings…</td>
<td>Field notes are organised and summarised with the learner and a narrative In Training 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.</td>
<td>A progress review document and individual learning plan is uploaded onto an ePortfolio 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)</td>
<td>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.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P11: Learners have recurrent learning meetings with (faculty) mentors/coaches using a self-analysis of all assessment data.</td>
<td>A faculty mentoring/coaching/academic tutoring system is in place</td>
<td>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 operating procedures. Evidence of engagement with this process also forms an important part of the progression criteria…</td>
<td>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.</td>
<td>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.</td>
<td>Students have recurrent (weekly) meetings with clinical skills faculty coaches in years 1 &amp; 2, but less so in years 3 &amp; 4. Coaches remain with their learners across the year and provide feedback on clinical skills development and help students plan future learning.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA seeks to gradually increase the learner’s agency and accountability for their own learning leading them to self-tailored support individual learning priorities.</td>
<td>Students are encouraged or required to reflect on feedback (narrative/real-time/written) in self-directed learning</td>
<td>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.</td>
<td>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.</td>
<td>Students review electronic portfolio in advance of tutor meetings and reflect on achievement of competencies and personal learning plan.</td>
<td>This is much harder for us to implement in years 3 &amp; 4 when our students are increasingly dispersed, and our course becomes more modularised.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

UG: undergraduate; PG: post graduate; NA: North America; EUR: Europe; AUS: Australia.
We still run our programme. Learning drives the programme, and our assessment philosophy is based on effective feedback, triangulation 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.

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 termly 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 by 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 often see low-stakes as high-stakes.

P5: The method chosen should depend on the educational justification for using that method.

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.

P6: The distinction between summative and formative is replaced by a continuum of stakes.

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)

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. 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.

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.

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 often see low-stakes as high-stakes.
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.
For most students, who have modalities, no deliberations are held in exam boards. For students about whom doubts are flagged in low-stakes assessments, assessments are low-stakes until we have enough data. Two assessors independently judge learner performance by triangulation of quantitative and qualitative assessment data in a digital portfolio following a defined rubric. 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 overview (partly customised, e.g. show all narrative feedback for the competency).

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 low-stakes flags are very predictive of learners who will be in difficulty on high-stakes review.

We examine data from multiple teaching groups (three hospitals with largely non-overlapping rotations over three years). 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 all assessment results and entries in a portfolio.

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. 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 overview (partly customised, e.g. show all narrative feedback for the competency).

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 throughout 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.
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; Lobst 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, FRCPG, 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


