An international study on the implementation of programmatic assessment: Understanding challenges and exploring solutions

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

MEDICAL TEACHER

Taylor & Francis
Taylor & Francis Group

Check for updates

An international study on the implementation of programmatic assessment: Understanding challenges and exploring solutions

Dario Torre[a] , Lambert Schuwirth[b] , Cees Van der Vleuten[c] and Sylvia Heeneman[d]

[a]Director of Assessment, and Professor of Medicine, University of Central Florida College of Medicine, Orlando, FL, USA; [b]College of Medicine and Public Health, Flinders University, Adelaide, Australia; [c]Department of Educational Development and Research, School of Health Profession Education, Maastricht University, Maastricht, The Netherlands; [d]Department of Pathology, School Health Profession Education, Cardiovascular Research Institute Maastricht, Maastricht University, Maastricht, The Netherlands

ABSTRACT

Introduction: Programmatic assessment is an approach to assessment aimed at optimizing the learning and decision function of assessment. It involves a set of key principles and ground rules that are important for its design and implementation. However, despite its intuitive appeal, its implementation remains a challenge. The purpose of this paper is to gain a better understanding of the factors that affect the implementation process of programmatic assessment and how specific implementation challenges are managed across different programs.

Methods: An explanatory multiple case (collective) approach was used for this study. We identified 6 medical programs that had implemented programmatic assessment with variation regarding health profession disciplines, level of education and geographic location. We conducted interviews with a key faculty member from each of the programs and analyzed the data using inductive thematic analysis.

Results: We identified two major factors in managing the challenges and complexity of the implementation process: knowledge brokers and a strategic opportunistic approach. Knowledge brokers were the people who drove and designed the implementation process acting by translating evidence into practice allowing for real-time management of the complex processes of implementation. These knowledge brokers used a 'strategic opportunistic' or agile approach to recognize new opportunities, secure leadership support, adapt to the context and take advantage of the unexpected. Engaging in an overall curriculum reform process was a critical factor for a successful implementation of programmatic assessment.

Discussion: The study contributes to the understanding of the intricacies of implementation processes of programmatic assessment across different institutions. Managing opportunities, adaptive planning, awareness of context, were all critical aspects of thinking strategically and opportunistically in the implementation of programmatic assessment. Future research is needed to provide a more in-depth understanding of values and beliefs that underpin the assessment culture of an organization, and how such values may affect implementation.

Introduction

Programmatic assessment is an approach to assessment aimed at optimizing the learning and decision function of assessment (Van Der Vleuten and Schuwirth 2005). Central to programmatic assessment is that it aims to combine or triangulate information in the most meaningful way, typically by combining parts of different assessments on the basis of their content or the way that combination meaningfully informs the learner of his or her achievements and progress in competency domains (Schuwirth and Van der Vleuten 2011; Van Der Vleuten et al. 2015). For this, it requires a mix of assessment methods and is driven by a continuous and longitudinal integration of assessment of and for learning. A programmatic assessment approach is therefore a set of key principles and ground rules that are important for an effective implementation of this approach rather than a method in itself (Heeneman et al. 2021). These principles ensure that every assessment is meaningful to learning, not purely decision oriented. Decisions are

Practice points

- A programmatic assessment approach includes a set of key principles and ground rules that are important for an effective design; however, its implementation remains a challenge.
- There were two major factors in managing the challenges and complexity of the implementation process: core implementers or knowledge brokers and a strategic opportunistic approach.
- Knowledge brokers used an agile and 'strategic opportunistic' approach to recognize and exploit opportunities, adapt to context, manage and take advantage of the unexpected to achieve implementation goals.
- The study contributes to the understanding of the intricacies of implementation processes of programmatic assessment, across different international health profession education institutions.
- Future research is needed to explore values and beliefs that underpin the assessment culture of an organization.

taken on the aggregated data, usually in the form of competence committees.

For most educators, programmatic assessment is conceptually attractive, because it provides an answer to assessment in constructivist models of learning and competency-based learning (Van der Vleuten and Schuwirth 2019). Also, there is evidence that programmatic assessment works and has a positive impact on the quality and outcomes of learning (Schut et al. 2021). However, despite its intuitive appeal, the implementation of programmatic assessment remains a challenge. There is still a dearth of knowledge on implementation of novel assessment systems in different specialties and institutions; further, potential adaptation strategies to overcome implementation barriers remain understudied (Anderson et al. 2021). Programmatic assessment involves a fundamentally different mindset from the classic summative paradigm (Harrison et al. 2017); a shift from an exclusive focus on passing and failing to the realization of a feedback and learning from assessment culture (Watling and Ginsburg 2019). Such an educational change contravenes the deep beliefs and culture of an organization—so-called paradigm. This is different from more superficial changes (e.g. from written to computer-based assessment) which are generally more straightforward and easier to complete (McGaghie 2015). In contrast, more fundamental changes in learning and assessment systems have been shown to take longer and are encountered with more resistance. An example of such a fundamental mind shift is the introduction of Problem-Based Learning (Dolmans et al. 2005) compared to traditional lecture-based education. Similarly, the advent of competency based medical education has required a fundamental change in the educational approach (Frank et al. 2010, Carraccio et al. 2002) For the implementation of these learning and assessment systems, a number of factors have been described to influence (successful) implementation, such as sufficient student and faculty member training, buy-in from stakeholders, the perceptions of students and teachers of theoretical concepts, affinity with the educational reform, and the availability of time and money (Albanese and Mitchell 1993; Harris et al. 2010; Hawkins et al. 2015; Iobst and Holmboe 2020).

The difficulties of implementations of fundamentally different learning and assessment systems illustrate that education is a complex phenomenon which is highly depending on the interaction between learners and knowledgeable others (Rosas 2017). Because this interaction and communication play a central role, it is logical that education and the meaning making in assessment are highly context and culture specific. Specific rules, regulations and legislation around education and assessment—also around secondary and postgraduate education—impact on how programmatic assessment or any other strategy for assessment for learning can be implemented. Culture and the associated philosophy of learning can thus have an enormous impact on the success of an implementation (Boyd et al. 2018) because they facilitate or limit the affordances (what the implemented approach allows/disallows the stakeholders to do) the organization will build into the design, and the effectivities (what the stakeholders are able/unable to use) the stakeholders can develop. One cannot just take a successful education/

assessment strategy from one place and 'copy' it into another (Waterval et al. 2015). Also, educational organizations have a high degree of complexity, and in a complex system, actions are interconnected, the relationship among parts is often more important than the parts themselves. Individuals and systems behave adaptively, processes are nonlinear, unpredictable, and may co-evolve (Waldrop and Gleick 1992).

Two main theoretical frameworks support our study: Institutional Logics (Thornton and Ocasio 2008; Gordon and Cleland 2021) and the Consolidated Framework for Implementation Research (CFIR) (Damschroder et al. 2009). In order to make sense of the complexity of the change process, we used the institutional logics perspective as one of our theoretical lenses. Institutional logics explores the interrelationship between individuals, organization and context arguing that every institution has a central logic, a socially constructed set of values and beliefs that influence individual and organizational behaviors. Multiple logics may coexist or compete in the same institution, one may be dominant over the other, and their impact on individuals and organizations will vary. A set of institutional logics, despite its inherent contradictions, provides individuals with agency and can shape change in human and organizational behaviors. For example, assessment may be influenced by the logics of teachers providing the assessments and by those of students involved in the process. These logics may be in contradiction with one another yet differences between logics may lead to new understandings and meaningful change.

The CFIR (Damschroder et al. 2009) is a theoretical lens to examine the specific factors and circumstances that surround the implementation across different contexts. The CFIR is comprised of five interconnected domains: intervention, inner and outer settings, individuals and process. The intervention is that what is being implemented (e.g. programmatic assessment) which may be internally or externally developed and has a core component (essential elements) and an adaptable periphery (elements that can be modified); the inner and outer setting, often interrelated, which include the structural, socio-political and cultural context within and outside the organization; the individuals who are involved in the implementation; the process with its subprocesses such as planning, engaging, executing that may be carried out in a linear and nonlinear fashion.

The implementation of programmatic assessment within an educational system has to deal with multiple and divergent perspectives, partial solutions and new strategies that require a full understanding of the relationality of people and processes (Ang 2011). Therefore, a deeper understanding of change management and implementation from the perspective of a complex adaptive system is critical for the successful implementation of a context-specific and culturally bound paradigm like programmatic assessment. This is important because before programmatic assessment is to drive learning and to exert its positive educational effects, it has to be fully and successfully implemented. Therefore, knowledge and further understanding of adaptation strategies as to how concepts of programmatic assessment are ultimately translated into practice, are of critical importance. Fortunately, programmatic assessment has been implemented in a number of

undergraduate (Wilkinson et al. 2011; Driessen et al. 2012; Bok et al. 2013; Roberts et al. 2014; Heeneman et al. 2017; Jamieson et al. 2017) and graduate educational programs (Chan and Sherbino 2015). The purpose of this paper is to learn from the successes and challenges of the implementations to gain a better understanding of the factors that affected the implementation process of programmatic assessment and how specific implementation challenges were managed across different programs.

The research questions guiding our inquiry were the following:

1. What were the reasons for changing to a programmatic assessment approach?
2. What were the strategies to manage assessment change and mitigate challenges encountered in the implementation of programmatic assessment across different contexts?

Methods

Design

An explanatory multiple case (collective) study approach was used for this study (Stake, 1995; Yin 2009; Cleland et al. 2021). The unit of analysis (the case) was each program that had implemented programmatic assessment and the purpose (Harrison et al. 2017) of the case was to examine how individuals, processes, and context interacted to implement programmatic assessment, and to explore different enablers, challenges between and across cases to gain a better understanding of the implementation process of programmatic assessment. Our ontological approach was aligned with Stake's constructivist orientation in which the reality is viewed as a 'universe of integrated interpretations' (Stake, 1995, p. 100). We took an interpretative position, taking on the role of interpreters or gatherers of interpretations of the experiences of the constructed implementation realities of the different programs (Stake, 1995). Further, within a constructivist paradigm, we assumed a relativist ontology position, considering that multiple different realities are present among informants and people interpretations vary and may not be of equal value.

Our epistemological stance follows Stake's (1995) constructivist perspective, which allows for a holistic understanding of a phenomenon (implementation of, or an assessment change to programmatic assessment) from the

perspective of those involved, in relation to the particular context in which it occurred.

Case selection and participants

Cases were purposefully selected because they could provide an understanding of the issues and insight into the implementation of programmatic assessment (Stake, 1995). Criterion based sampling was used to identify cases (programs) with the richest information, from which we could 'maximize what we can learn' (Stake, 1995, p. 4) and understand the key aspects of the programmatic assessment implementation process.

In order to focus the data collection, the cases were bounded (Stake, 1995) by the following parameters: (1) programs and individuals who participated in the 2020 Ottawa consensus statement on programmatic assessment (Heeneman et al. 2021; Torre et al. 2021); (2) programs that implemented programmatic assessment in either undergraduate or post-graduate across different health profession disciplines; (3) University based programs from different geographic areas thus different sociocultural contexts, that implemented programmatic assessment for at least 1 year. None of the authors were involved in the study as cases.

Data collection

We identified 6 programs (see Table 1) with maximum variation regarding health profession disciplines, level of education and geographic location. We contacted one individual at each of the 6 programs to participate in the study. These individuals were health professionals from different health profession education disciplines (medical, dentistry, veterinary), worked in undergraduate and graduate health profession education, and were also part of the 2020 Ottawa consensus group on programmatic assessment (Heeneman et al. 2021; Torre et al. 2021). They were recommended by at least two of the authors (CPM, LS) as being knowledgeable and as having played a key role in the implementation of programmatic assessment at their institution (Dannefer et al. 2012; Bok et al. 2013, Wilkinson and Tweed 2018). Those individuals were health profession practitioners and educators, either directors of assessment or individuals responsible for the assessment program at time of implementation, thus were deemed to be key informants for each case.

Table 1. Characteristic of the cases.

Case	Discipline	Location/university	Undergraduate/ postgraduate	Year of first implementation:	Number of students	Length of the program
1	Medicine	USA	Undergraduate	2004, in pre-clinical and clinical phase	32 students/ year	5 years
2	Medicine	New Zealand	Undergraduate	2006, iterative process, in pre-clinical and clinical phase	300 students/ year	6 years
3	Medicine	Canada	Undergraduate	2016, in pre-clinical and clinical phase	260 students/ year	4 years
4	Dentistry	UK	Undergraduate	2019, in pre-clinical and clinical phase	377 students/ year	5 years
5	Medicine	Canada	Postgraduate, General Practice	2004, iterative process, residency training	136 residents, 9 training sites	3 years
6	Veterinary	Netherlands	Undergraduate	2010, in clinical phase, i.e. the final 3 years of undergraduate program	218 students /year	3 years

The principal and senior investigator (DT, SH) developed an interview guide based on previous literature (Heeneman et al. 2015; Van Der Vleuten et al. 2015; Timmerman and Dijkstra 2017; Wilkinson and Tweed 2018; Ross et al. 2021) and on the recent report from the 2020 Ottawa consensus statement on programmatic assessment (Heeneman et al. 2021, Torre et al. 2021). The guide included two sets of interview questions. The first set focused on the reasons for the implementation of programmatic assessment. The second set of questions focused on the initial steps taken in the implementation process, what facilitated the process and how challenges encountered were managed to achieve a successful implementation. We chose interviews because they allowed us, in partnership with the participants, to garner their interpretation and meaning of the implementation process. As stated by Stake (1995), 'the interview is the main road to multiple realities' (1995, p. 64)

One author (SH) contacted all participants and conducted all the interviews. Interviews were performed on Zoom using an encryption feature. Digital recordings were transcribed by a specialized agency.

Ethical considerations

Participation was voluntary and participants were ensured of confidentiality and signed an informed consent form. The ethical review board of the Dutch Association for Medical Education approved this study (approval NVMO-ERB-2020.3.4).

Data analysis

We used deductive thematic analysis of the transcripts based on the research questions and the theoretical lenses of CFIR and Institutional Logics frameworks. For this analysis, we adopted the six-step approach of thematic analysis outlined by Braun and Clarke (2006). SH, LS and DT reviewed independently two initial transcripts to become familiar with the data (step 1) and generated initial codes. DT, SH, LS met to reconcile discrepancies and to generate a preliminary codebook. They then individually coded two additional transcripts using this codebook and made further refinements to a finalized codebook. Finally, DT, SH, LS coded the remainder of the transcripts (step 2). Throughout the entire analysis process, SH, LS, and DT met regularly to share their understanding of codes, discuss the meaning of new codes, and identify potential connections among codes. The identified codes were sorted and combined into potential themes, collating coded data within and across cases, developing a thematic map (step 3). SH, LS and DT reviewed and refined previously generated themes by a recursive analysis of codes and themes. Themes were eventually modified to better capture the coded data (step 4). Themes were defined, and named, organizing them into a coherent narrative to provide a unique insight into the research questions (step 5).

A draft of the manuscript was written and shared with the remaining author (CvdV) who reviewed the analysis and provided additional insight and interpretations (step 6). We used ATLAS ti software program for our analysis (ATLAS ti 2020).

Reflexivity

DT is a practicing physician and a trained medical education researcher. SH is a biomedical PhD with training in educational research. LS is a physician, with expertise in medical education and an expert educational researcher in programmatic assessment. CvdV is a psychologist by training, and one of the leading researchers on programmatic assessment. We realize that our own beliefs, and values influenced our analytic process and recognize our subjectivity in the interpretive process. In accordance with a constructivist epistemology, we were aware that our assumptions and views informed and shaped the knowledge generated by our research experiences (Crabtree 1999). Therefore, in order to balance our perspectives and attend to our own biases, an audit trail was kept of discussions and decisions made, which was regularly revisited. All members of the team provided input throughout the analysis process engaging in iterative discussions about the data and insights gained. Given our own expertise and experiences in the theory and practice of programmatic assessment, the audit trail, iterations and discussions enabled us to keep a critical stance on the data, and unexpected insights.

Rigor was also ensured by investigator triangulation (Stake, 1995). Three researchers (DT, SH, LS) reviewed and analyzed data from all interview transcripts and discussed alternative interpretations. Data was also presented and discussed with a fourth researcher (CvdV).

Results

A total of 6 interviews with six faculty members from different programs worldwide were conducted and analyzed. Table 1 provides demographic characteristics of the 6 programs.

We will present our findings in two parts. We will first describe participants' general perspective on why change in each program was needed (Table 2); then the two factors identified as critical in managing the challenges and complexity of the implementation of programmatic assessment are presented: people and the use of strategic opportunism. Salient quotes are represented to illustrate the main themes.

The perception of a need for change

From the data, it became clear that the main driver was not so much some reason for change but the perception of the need for change. For example, external changes such as competency-based frameworks, new national competency-based frameworks or accreditation requirements in themselves may not have been sufficient to drive the change towards programmatic assessment. Essential is the way in which these drivers were seen as forces towards the implementation of programmatic assessment. This is important because the types of drivers were quite varied, and the main common feature was how to all led to a perceived need to achieve a specific outcome. Consequently, the desired outcome varied by context as well, and they were always a combination of trying to align internal processes with external developments (see Table 2).

Table 2. Overview of the reason for change to programmatic assessment for each case.

Case	Reason for change	Quote
1	*Internal purpose*: fostering Self-Directness Learning and clinical expertise, developing a new learning environment	'we tried to develop a curriculum and assessment system that would foster self-directed learning, a broad-based clinical expertise and not be the medical school that everyone else went to [laughs]. To have a positive and encouraging learning environment where we would produce future reflective practitioners. So it wasn't necessarily a problem we were trying to solve, it was a learning environment we wanted to create'
2	*Internal purpose*: focus student outcome, end of year decision/ outcome not a surprise for student	'.. we wanted to make sure that we captured the attributes that were more than just the things captured on traditional tests. We wanted to make sure that we incorporated professionalism and not just knowledge and not just skills. That was one. The other big problem we were facing was the surprise that some students got at the end of the year when they suddenly found that they might be going to fail and indeed the surprise that some of the staff suddenly at the end of the year we would say oh, this person hasn't done very well this year and why didn't we pick it up earlier?'
3	*Internal purpose*: curriculum renewal, emerging evidence of assessment theory *External demand*: National Competency-based Medical Education movement	'..the reasons for both the curriculum and assessment renewal aligned with emerging evidence, articulations of societal need. Some of the themes around it were, well, both national and international bodies calling for curricula that are more integrated, and assessment as well, and more of a move to engendering metacognitive capacities in our learners. So reflective capacity and other metacognitive capacities, resilience, and I think more tailored learning and assessment, more individualised, recognising different paths of—different trajectories for people.'
4	*Internal purpose*: lack of assessment feedback reported by students' evaluations *External demand*: need of regulators to 'measure' outcomes	'Back in 2008, the big problem was that the general dental council who is our regulator was moving [..] to an outcome-based model called Prepared for Practice. So, Prepared for Practice presented quite a number of challenges because one of the things that they wanted us to measure, for example, was professionalism.' '..the other issue we had is we have a thing in the UK called the National Student Survey. In their final year students get asked a whole bank of questions, the school gets an overall ranking. We were doing really, really badly in the National Student Survey, I think overall there was about 40 to 45 per cent overall satisfaction. One of the bigger area of problems was assessment feedback.'
5	*External demand*: perceived national problem of residents not adequately prepared for practice,	'.. it was a national problem. It wasn't just a local problem. I was involved in a committee nationally for the College of Family Physicians of Canada. It was called the Working Group on the Certification Process. The college itself was getting a lot of complaints from special-interest groups—Diabetic Association, Rural Physicians in Canada—saying basically that our residents were not trained adequately … … …'
6	*Internal purpose*: curricular renewal (based on SWOT), need to document feedback better, identify and remediate underperforming students	'..we performed a SWOT analysis of our previous program. What we find out was that—well there were several reasons for changing the program. First, it was very difficult for us to follow-up on students over time. It was very difficult to follow-up on student's assessment results in a certain—within a certain clerkship and then follow it up to the next—to the following clinical clerkship. Students feedback—the feedback students get; all their assessments wasn't recorded. There was no database or no repository of assessment information. It was very difficult to see and to judge how students progressed over time. As a consequence, it was very difficult to judge if a student, especially the students that were performing below a certain threshold. It was very difficult to identify those students and to remediate those students of course as a consequence. We needed a system that allowed us to identify students—the progress students made over time, over different clinical rotations and not only within a certain course or within a certain clinical rotation but [..] the longitudinal aspect and documenting assessment results, so documenting feedback.'

The perception of this need could be initially internally driven, for example a certain educational need or to solve a perceived educational or assessment (design) problem in accordance with modern scientific literature and developments. Alternatively, it could be initially dictated new national competency-based frameworks or accreditation requirements requiring adaptation of processes.

People's roles

Logically, people were the drivers and designers of the implementation process because as stated before, people are the ones who perceive the driving forces and construct the optimal way to align internal processes to external drivers. People's perceptions played a key role in the implementation process acting as core implementation guidance, and in in supporting both seeking and providing leadership buy in. These core implementers surfaced as 'knowledge brokers,' who using their networking capabilities, expertise and assessment literacy, were able to begin, design and manage the complex process of implementation. Knowledge brokers acting upon internal and external drivers for change, moved strategically within the system to assure the support of leadership.

Knowledge brokers

The majority of knowledge brokers took a primary role in initiating the implementation process, and took a lead position to affect change as being the:

'… lead for the implementation of programmatic assessment in our first two years'… (Case 3) or served as 'chair of our assessment committee'. (Case 2)

Knowledge brokers purposefully formed strategic alliances internally, with influential faculty members, early adopters, within the program by identifying:

…local champions. …key role models, people in key positions, … … who were respected who also shared those views [of programmatic assessment] …. (Case 2)

They managed to establish ongoing collaborations and strategic alliances with other medical school programs that were implementing programmatic assessment at the same time:

We were both doing this at the same time and many things were quite aligned with how we were approaching. So we

were able to be in regular touch as well and that helped. (Case 3)

Leadership

The knowledge brokers were strategic in seeking and securing leadership buy-in and support, such as from, deans, vice deans, program directors etc. Leadership support was critical for the implementation and sustainability of programmatic assessment in all programs. It was a way to steer processes using a combination of intrinsic meaning making—through communication and staff education—and more managerial drivers. The role of leadership was somewhat different across programs, yet their support was pivotal, particularly at the beginning of the process when resistance to change was high and challenges were greater:

> They [leadership] always had our back, they always supported us, also over the years, especially within the early years because there were some bad evaluations . from students, from teachers in the first year. (Case 6)

Strategic opportunism

Implementation of programmatic assessment is a process that occurs in a complex system. In complex systems, the parts are intertwingled, connected by non-linear, recursive relationships such that one agent changes the context for another agent. We identified an agile and strategic opportunistic approach performed by a number of knowledge brokers. Rather than working from a static project planning they acted upon internal and external need for change, designed, adapted and sustained the implementation process while dealing with its challenges. They were able to recognize new opportunities and take advantage of the unexpected, while attending to the current needs and daily occurrences of their job. Adaptation to context, promoting assessment literacy among faculty and students, fostering early sustainability, and developing a strategic alignment between assessment and curriculum redesign were identified as multiple parallel processes knowledge brokers leveraged to navigate complexity. But the knowledge brokers were not just agile and adaptable, but also goal oriented. They had clear conceptualisations of the purpose of the change.

> So. it wasn't necessarily a problem we were trying to solve; it was a learning environment we wanted to create. (Case 1)

> I think there were two things. I think we wanted to make sure that we captured the attributes that were more than just the things captured on traditional tests. We wanted to make sure that we incorporated professionalism and not just knowledge and not just skills. (Case 2)

So, with the term 'strategic opportunism' we mean strategic in having a clear goal or outcome in mind but being situationally aware and agile in the ways to achieve those goals. Obviously, strategic opportunism requires expertise to enable a constant, valid translation of published evidence and others' experiences to the local context.

Adaptation to context

Adapting to and making compromise in the setting of different contexts was, therefore, critical for implementation purposes. A certain degree of uncertainty and ongoing tensions were seen that were inevitable in the setting of change within a complex system.

This was especially important in mitigating challenges and reach a successful implementation and at the same time keeping true to the principles of programmatic assessment. Although not all tensions could be solved, some degree of compromise was necessary.

> Keeping in mind what your end goal is with this new approach [programmatic assessment] . then sometimes you have to make a compromise to at least be able to implement a lot of …. our programmatic approach and the compromise was that, … we allow some clerkships to have a few knowledge tests within the program … … … … and because we allowed these knowledge testing in these clerkships, they were also implementing the other [principles] … .. (Case 6)

Similarly, when other educational institutions would require specific outcomes such as grade point average or class ranking for admission to postgraduate studies, tensions would emerge that required continuing negotiation and adaptive approaches.

> . we've tended to sidestep that [GPA or class ranking], and just say well … we don't do it that way, that's your problem not ours, and we sort of got away with that for quite a while, but just more recently it's coming back. (Case 1)

Rules and regulation in the local and national educational context also played a role, yet their impact on implementation was different across contexts. In program 2, national regulations and accreditation standards did not constitute a barrier and seem to foster innovative practices.

> We don't have a national exam so there's nothing external to what the university

> does that is impacting on assessment directly. …. They [accreditation bodies] just want to know is it effective, is it working, is it consistent with other things you're doing? You could argue even that they were facilitators, both the absence of … a national exam and the permissive accreditation standards you could argue actually helped. (Case 2).

In program 3, the implementation of programmatic assessment was fully aligned with accreditation and regulation processes which thus promoted its implementation rather than hamper it:

> Our national accreditation framework, for the most part, I would say did not pose a huge challenge to what we wanted to do. If anything, I think in most of the standards, what we've done with programmatic assessment just helps our case …. (Case 3)

Implementation processes were clearly different across all six programs because they were adaptive to each context, culture, historical moment and current structure. In some programs, the process started off with the development of a central oversight or progress committee that '*went through the process of designing all of the assessments.and was critical*' to accomplish a successful implementation *(Case 1, 2, 6)*, in others, people '*took a project management approach with external project management consultants that … were on board for probably maybe a year and a half or two years*' *(Case 3, 4)*.

> But even with the project management approach agile processes took place.

> Basically, from day 1 the planning of the curriculum and the assessment were done alongside each other, with each informing the other, which I think was very important. (Case 3)

Promoting assessment literacy among faculty and students

To promote transformation, purposeful and ongoing faculty development and student support programs were developed and conducted. Most often, faculty development was performed right at the beginning of implementation and continued throughout the implementation process.

> . from the outset, building up was the faculty development lens was a lens that was worn also right from the design phase all the way through implementation. So it wasn't like, oh, we built it and then how do we develop the faculty?. (Case 3)

Many programs emphasized the importance of providing students' with knowledge and understanding of the assessment process, clarifying expectations, explaining the importance of seeking feedback within the programmatic assessment framework.

> The role of the student is important. We also invest a lot in assessment literacy with students. . students they demand …. that they [faculty] provide them with meaningful feedback ….. (Case 6)

In some programs knowledge brokers capitalized on the opportunity to use evidence from assessment theory and practice to foster assessment literacy among faculty. They developed progress/assessment committees that not only served the purpose of reviewing and assessing students' progress and also served as a forum for faculty to share their views about assessment, eventually creating consensus among faculty and fostering a transformative process. This was based on their knowledge of evidence from assessment theory and practice.

> There's the joint decision-making principle, the decisions are made by a group of people. It's not just a principle of programmatic assessments but I actually think that people really like doing it because they feel supported ….—one of the other principles is [about] the stakes of your decision. You need lots of pieces of information before you make a high stakes decision. People like that as well because they all realize that their assessments are all a bit imperfect but together they're better. (Case 2)

Early sustainability

It is important to recognise that change is not an event but an ongoing process and when issues or challenges are identified, it is important to provide a response. In all programs, sustainability was not an afterthought but an integral part of the process and a strategy to provide ongoing attention and promote a culture of improvement. Knowledge brokers had a key understanding of what was needed for sustainability, such as resources, creating new roles/or positions, and training new incoming teachers about the principles of programmatic assessment.

> I think there's the creators and then there's the sustainers and I think it requires more resources and commitment to sustain a program than it does to create a program. [.]. I was on the creating part, but it's always working to try to make sure that your monitoring parts of the system and trying to prove that and advocating for resources because a curriculum is a living organism. (Case 1)

Alignment of assessment and curriculum redesign

Strategic choices were also made at the start of the implementation process. Although programmatic assessment is thought of and presented often as an assessment model, in all programs change entailed a combined and aligned (re)design of education and assessment. The alchemy of assessment and curriculum was clearly present. The redesign of assessment in parallel with curriculum transformation ultimately facilitated the implementation of programmatic assessment as the assessment model.

> Basically, from day 1 the planning of the curriculum and the assessment were done alongside each other, with each informing the other, which I think was very important. (Case 3).

> . we were in the lucky situation that we had a complete curriculum renewal … That was an excellent opportunity for us to also reform our assessment strategies. That was very lucky for us because I think implementing programmatic assessment within a program that is already running and that's a program that will not change. I think that's more difficult but, in our case, the entire curriculum was reformed. That was an excellent opportunity for us to also reform our assessment strategies. (Case 6)

It was clearly recognized that curriculum and assessment are closely intertwined and that, the implementation of one cannot succeed without changes in the other.

Discussion

The study contributes to the understanding of the intricacies of the implementation process of an assessment system, programmatic assessment, across a number of different international health profession education institutions. The reasons for change were context-specific and were either responding to an internally perceived purpose and/or to an external demand. Two major factors were identified eminent to manage change and mitigate challenges. Firstly, people, who were knowledgeable and familiar with the principles of programmatic assessment, and secondly, a strategic opportunistic approach that seized new opportunities to promote change while striking a balance between short term goals and long-term direction. Additional commonalities were found across programs; Agile and strategic adaptation to different contexts was an essential skill for individuals to implement programmatic assessment. The symbiotic relationship between curriculum change and programmatic assessment seemed to be of critical importance. Programmatic assessment principles were negotiated among stakeholders, reshaped, and adjusted to fit social and institutional contexts without losing their core elements.

Implementation of major changes in an educational context is a complex adaptive process, often unpredictable yet manageable (Plsek and Wilson 2001; Sanger and Giddings 2012). It is predicated on multiple interactions between various types of stakeholders, with various beliefs, worldviews and assumptions about the reality, their organization and cultural context (Jippes et al. 2013). Complexity theory suggests that change emerges from the dynamic complexity of the interactions among structures and people in the environment (Mason 2009). In our study, leadership played an important 'enabling' function in the implementation process of programmatic assessment. This is described in complexity leadership theory, a management theory, in which enabling leadership facilitates connections between the traditional hierarchical, regulation-driven leadership and the more adaptive, and innovative 'entrepreneurial' leadership of the

organization (Gordon and Cleland 2021) resulting in organizational change.

Systematic exploitation of opportunities, adaptive planning, awareness of contextual changes, possessing the flexibility to manage the unexpected, were all critical aspects of thinking strategically and opportunistically to manage the implementation of programmatic assessment in a complex environment. This complexity was further increased by the fact that successful implementation of approaches such as programmatic assessment are not just one process. The successful implementations in the cases presented were all a combination of adaptive, interconnected parallel processes: first, there was the 'pure' implementation. Changes were made to the assessment model and were driven by a wide array of internal or external needs. Second, adjacent to the pure implementation, existed a process of expertise management involving staff development on assessment literacy, induction into the nature and principles of programmatic assessment, students' education about a new assessment culture, and individuals' capacity to adapt and respond to a dynamic, unpredictable environment. The third process relates to a scaffolding management. Processes are in place to ensure that eventualities can promptly be dealt with, and that continuous ongoing support is provided to assist and sustain the program while communicating with all stakeholders and finally, a process of evaluation and other meta-activities; especially in an ever-changing staffing context, it becomes important to convince new stakeholders, whether they are new faculty, students or other health professionals to continue their support of the assessment model and its implementation.

Successful change begins and ends with people (Hord and Roussin 2013). If it is true that organizations adopt change, individuals implement them (Hall and Hord 2001). People who are successful agents of change, are resilient, tolerate ambiguity and adapt to uncertainty (Woodruff 2019). In order to navigate a complex adaptive system and ensure that evidence is successfully translated to local context and that the parallel processes are aligned in such a way that they can act synergistically, people or individuals with specific expertise are needed within the organization. As these people have to play multiple roles—that of expert, manager and sometimes even project owner—they can best be described as 'knowledge brokers.' Knowledge brokers are able to capitalize on serendipity, align different sources of power, adapt to everchanging context, while managing to find solutions to complex problems. Knowledge brokers resemble what Wenger-Trayner and Wenger-Trayner (2014) calls 'system conveners.' System conveners are people who seek and create new learning partnerships, build connections among stakeholders, spot opportunities to bring people together in order to reconfigure complex landscapes and create lasting change (Wenger-Trayner and Wenger-Trayner 2014). Further, in accordance with institutional logics approach, institutional entrepreneurs were change agents that played a critical role in exposing the contradictions of dominant institutional logics to bring about change (Thornton and Ocasio 2008).

Health profession education systems are highly complex systems (Mennin 2013). Such complexity makes the implementation process unpredictable and therefore not amenable to a stepwise project plan or script. The unpredictability is not infinite however, it is a bounded unpredictability. It will be clear in the process that certain pathways to solving the implementation problems are unviable or 'out of bounds' but within the problem space, multiple solutions can be developed and used. Ultimately, each educational system will likely plan and design different pathways, and processes to implement the principles of programmatic assessment, yet it will be critical to adapt the implementation to the unique structural, cultural, and political context of the organization in which it will take place (Ross et al. 2021).

Previous evidence on managing curriculum reforms, has shown that the alignment of instructional methods, assessment, students and faculty needs is a critical aspect for a successful curriculum implementation (Kulasegaram et al. 2018). In our study, we found that the implementation of programmatic assessment was closely intertwined with curriculum transformation. We argue that, given the symbiotic and interdependent relationship between programmatic assessment and curriculum, the implementation of a programmatic assessment model would be most meaningful and effective in the setting of a curriculum-wide transformation process.

This study has limitations. First, in this study no data was collected over time, exploring how change occurred longitudinally. However, the scope of our study was not to assess change over time but to gain an understanding of the factors that led to a successful implementation process. Second, we do not contend that these programs represent all the institutions that have implemented programmatic assessment. We realize there may be other institutions that implemented programmatic assessment, and some may have had a different process and a negative or unsuccessful outcome. However, in accordance with a Stakian approach (1995) we purposefully selected programs from which we could learn the most about the implementation of programmatic assessment in accordance with its principles. Further research is needed to ascertain whether the factors we identified can be transferred to different contexts. This being said, parallels can be drawn between the feature of the implementations we identified and those known in the domain of project management. Here too, similar approaches, so called agile techniques have become popular and effective (Serrador and Pinto 2015). Third, there might have been individuals in the implementation process whose voices we did not capture; however, given their role in the process, respondents were the richest information source to maximize what could be learned from each case. Finally, the cases we selected did not include educational programs from Asia, Africa, South America, or other areas of the world in which a different culture and context may have had a different impact or outcome on the implementation of programmatic assessment. It would be interesting to know how culture shapes the implementation of programmatic assessment, just like in self-directed learning (Frambach et al. 2012). it is important to acknowledge that as medical educators, we should strive for equitable, context-sensitive and locally driven approaches to programmatic assessment (Gosselin et al. 2016). In conclusion, acknowledging the role of motivated,

knowledgeable and resilient people who strategically act on opportunities within and outside the organization is crucial for anyone who engages in the implementation of programmatic assessment. The implementation process is non-linear and has multiple components that interact with each other. Despite the inherent complexity of the system, people drive the process; they form alliances to gain leadership support, find ways to compromise with rules and regulations, stay flexible to recognize new opportunities in a constant adaptation mode to face the challenges of a complex and dynamic context. Future research is needed to provide a more in depth understanding of the values and beliefs that underpin the assessment culture of an organization, and how such values may affect the implementation of programmatic assessment. Further research is also needed to further explore factors that may prevent or derail implementation in specific settings and contexts. Finally, additional inquiry may delve into on how implementation practices may affect learning behaviors.

Disclosure statement

Glossary

Programmatic assessment: Is an approach to assessment aimed at optimizing the learning and decision function of assessment.

Funding

Notes on contributors

Dario Torre, MD, MPH, PHD, is professor of Medicine and Director of programs of assessment at University of Central Florida College of Medicine.

Lambert Schuwirth, MD, PHD, is professor of medicine at Flinders University College of Medicine and Public Health, Adelaide, Australia.

Cees Van der Vleuten, PHD, is Professor of Education Department of Educational Development and Research, School of Health Profession Education, Maastricht University, The Netherlands.

Sylvia Heeneman, PHD, is Professor of Education Department of Pathology, School Health Profession Education, Cardiovascular Research Institute Maastricht, Maastricht University, The Netherlands.

ORCID

Dario Torre http://orcid.org/0000-0002-4924-4888
Lambert Schuwirth http://orcid.org/0000-0002-6279-5158
Cees Van der Vleuten http://orcid.org/0000-0001-6802-3119

References

Albanese MA, Mitchell S. 1993. Problem-based learning: a review of literature on its outcomes and implementation issues. Acad Med. 68(1):52–52.

Anderson HL, Kurtz J, West DC. 2021. Implementation and use of workplace-based assessment in clinical learning environments: a scoping review. Acad Med. 96(11S):S164–S174.

Ang I. 2011. Navigating complexity: From cultural critique to cultural intelligence. Continuum. 25(6):779–794.

ATLAS ti. version 9.1 Qualitative Data Analysis. 2020.

Bok HG, Teunissen PW, Favier RP, Rietbroek NJ, Theyse LF, Brommer H, Haarhuis JC, van Beukelen P, van der Vleuten CP, Jaarsma DA. 2013. Programmatic assessment of competency-based workplace learning: when theory meets practice. BMC Med Educ. 13(1):1–9.

Boyd VA, Whitehead CR, Thille P, Ginsburg S, Brydges R, Kuper A. 2018. Competency-based medical education: the discourse of infallibility. Med Educ. 52(1):45–57.

Braun V, Clarke V. 2006. Using thematic analysis in psychology. Qual Res Psychol. 3(2):77–101.

Carraccio C, Wolfsthal SD, Englander R, Ferentz K, Martin C. 2002. Shifting paradigms: from Flexner to competencies. Acad Med. 77(5): 361–367.

Chan T, Sherbino J, 2015. The McMaster Modular Assessment Program (McMAP): a theoretically grounded work-based assessment system for an emergency medicine residency program. Acad Med. 90(7): 900–905.

Cleland J, MacLeod A, Ellaway RH. 2021. The curious case of case study research. Med Educ. 55(10):1131–1141.

Crabtree BF. 1999. Doing qualitative research. London: SAGE.

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):50–15.

Dannefer EF, Bierer SB, Gladding SP. 2012. Evidence within a portfolio-based assessment program: what do medical students select to document their performance? Med Teach. 34(3):215–220.

Dolmans DH, De Grave W, Wolfhagen IH, Van Der Vleuten CP. 2005. Problem-based learning: future challenges for educational practice and research. Med Educ. 39(7):732–741.

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.

Frambach JM, Driessen EW, Chan LC, van der Vleuten CP. 2012. Rethinking the globalisation of problem-based learning: how culture challenges self-directed learning. Med Educ. 46(8):738–747.

Frank JR, Snell LS, Cate OT, Holmboe ES, Carraccio C, Swing SR, Harris P, Glasgow NJ, Campbell C, Dath D, et al. 2010. Competency-based medical education: theory to practice. Med Teach. 32(8):638–645.

Gordon L, Cleland JA. 2021. Change is never easy: how management theories can help operationalise change in medical education. Med Educ. 55(1):55–64.

Gosselin K, Norris JL, Ho M-J. 2016. Beyond homogenization discourse: reconsidering the cultural consequences of globalized medical education. Med Teach. 38(7):691–699.

Hall GH, Hord SM. 2001. Implementing change: patterns, principles, and potholes, 4th ed. Boston: Peason.

Harris P, Snell L, Talbot M, Harden RM. 2010. Competency-based medical education: implications for undergraduate programs. Med Teach. 32(8):646–650.

Harrison CJ, Könings KD, Schuwirth LW, Wass V, Van der Vleuten CP. 2017. Changing the culture of assessment: the dominance of the summative assessment paradigm. BMC Med Educ. 17(1):1–14.

Harrison H, Birks M, Franklin R, Mills J. 2017, January. Case study research: foundations and methodological orientations. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* (Vol. 18, No. 1).

Hawkins RE, Welcher CM, Holmboe ES, Kirk LM, Norcini JJ, Simons KB, Skochelak SE. 2015. Implementation of competency-based medical education: are we addressing the concerns and challenges? Med Educ. 49(11):1086–1102.

Heeneman S, de Jong LH, Dawson LJ, Wilkinson TJ, Ryan A, Tait GR, Rice N, Torre D, Freeman A, van der Vleuten CP. 2021. Ottawa 2020 consensus statement for programmatic assessment–1. Agreement on the principles. Med Teacher. 43(10):1–10.

Heeneman S, Oudkerk Pool A, Schuwirth LW, van der Vleuten CP, Driessen EW. 2015. The impact of programmatic assessment on student learning: theory versus practice. Med Educ. 49(5):487–498.

Heeneman S, Schut S, Donkers J, van der Vleuten C, Muijtjens A. 2017. Embedding of the progress test in an assessment program

designed according to the principles of programmatic assessment. Med Teacher. 39(1):44–52.

Hord SM, Roussin JL. 2013. Implementing change through learning: concerns-based concepts, tools, and strategies for guiding change. Thousand Oaks: Corwin Press.

Iobst WF, Holmboe ES. 2020. Programmatic assessment: the secret sauce of effective CBME implementation.

Jamieson J, Jenkins G, Beatty S, Palermo C. 2017. Designing programmes of assessment: a participatory approach. Med Teach. 39(11):1182–1188.

Jippes M, Driessen EW, Majoor GD, Gijselaers WH, Muijtjens AM, van der Vleuten CP. 2013. Impact of national context and culture on curriculum change: a case study. Med Teach. 35(8):661–670.

Kulasegaram K, Mylopoulos M, Tonin P, Bernstein S, Bryden P, Law M, Lazor J, Pittini R, Sockalingam S, Tait GR, et al. 2018. The alignment imperative in curriculum renewal. Med Teach. 40(5):443–448.

Mason M. 2009. Making educational development and change sustainable: insights from complexity theory. Int J Educ Dev. 29(2): 117–124.

McGaghie WC. 2015. Mastery learning: it is time for medical education to join the 21st century. Acad Med. 90(11):1438–1441.

Mennin SP. 2013. Health professions education: complexity, teaching, and learning. In Handbook of systems and complexity in health. New York, NY: Springer; p. 755–766.

Plsek PE, Wilson T. 2001. Complexity, leadership, and management in healthcare organisations. BMJ. 323(7315):746–749.

Roberts C, Shadbolt N, Clark T, Simpson P. 2014. The reliability and validity of a portfolio designed as a programmatic assessment of performance in an integrated clinical placement. BMC Med Educ. 14(1): 1–11.

Ross S, Hauer KE, Wycliffe-Jones K, Hall AK, Molgaard L, Richardson D, Oswald A, Bhanji F, ICBME Collaborators. 2021. Key considerations in planning and designing programmatic assessment in competency-based medical education. Medical Teacher. 43(7):758–764.

Rosas SR. 2017. Systems thinking and complexity: considerations for health promoting schools. Health Promot Int. 32(2):301–311.

Sanger M, Giddings MM. 2012. A simple approach to complexity theory. J Social Work Educ. 48(2):369–376.

Schut S, Maggio LA, Heeneman S, van Tartwijk J, van der Vleuten C, Driessen E. 2021. Where the rubber meets the road – an integrative review of programmatic assessment in health care professions education. Perspect Med Educ. 10(1):6–13.

Schuwirth LWT, Van der Vleuten CPM. 2011. Programmatic assessment: from assessment of learning to assessment for learning. Med Teach. 33(6):478–485.

Serrador P, Pinto JK. 2015. Does Agile work? – A quantitative analysis of agile project success. Int J Project Manage. 33(5):1040–1051.

Stakes RE. 1995. The art of case study research. Thousand Oaks, CA: SAGE.

Thornton PH, Ocasio W. 2008. Institutional logics. The Sage Handbook of Organizational Institutionalism. 840(2008):99–128.

Timmerman AA, Dijkstra J. 2017. A practical approach to programmatic assessment design. Adv Health Sci Educ Theory Pract. 22(5): 1169–1182.

Torre D, Rice NE, Ryan A, Bok H, Dawson LJ, Bierer B, Wilkinson TJ, Tait GR, Laughlin T, Veerapen K, et al. 2021. Ottawa 2020 consensus statements for programmatic assessment 2: implementation and practice. Medical teacher43(10):1149–1160.

Van Der Vleuten CP, Schuwirth LW. 2005. Assessing professional competence: from methods to programmes. Med Educ. 39(3):309–317.

Van der Vleuten CP, Schuwirth LW. 2019. Assessment in the context of problem-based learning. Advin Health Sci Edu. 24(5):903–904.

Van Der Vleuten CP, Schuwirth LWT, Driessen EW, Govaerts MJB, Heeneman S. 2015. Twelve tips for programmatic assessment. Med Teach. 37(7):641–646.

Waldrop MM, Gleick J. 1992. Complexity: the emerging science at the edge of order and chaos [m]. info. London: Viking.

Waterval DG, Frambach JM, Driessen EW, Scherpbier AJ. 2015. Copy but not paste: a literature review of crossborder curriculum partnerships. J Stud Int Educ. 19(1):65–85.

Watling CJ, Ginsburg S. 2019. Assessment, feedback and the alchemy of learning. Med Educ. 53(1):76–85.

Wenger-Trayner E, Wenger-Trayner B. 2014. Learning in a landscape of practice: a framework. In Learning in landscapes of practice. New York: Routledge; p. 13–29.

Wilkinson TJ, Tweed MJ. 2018. Deconstructing programmatic assessment. Adv Med Educ Pract. 9:191–197.

Wilkinson TJ, Tweed MJ, Egan TG, Ali AN, McKenzie JM, Moore M, Rudland JR. 2011. Joining the dots: conditional pass and programmatic assessment enhances recognition of problems with professionalism and factors hampering student progress. BMC Med Educ. 11(1):1–9.

Woodruff JN. 2019. Accounting for complexity in medical education: a model of adaptive behaviour in medicine. Med Educ. 53(9):861–873.

Yin RK. 2009. Case study research: design and methods (Vol. 5). Thousand Oaks: SAGE.