






RESEARCH ARTICLE

ASSESSMENT

Student perspectives on programmatic assessment in a large medical programme: A critical realist analysis

Chris Roberts¹  | Priya Khanna¹  | Jane Bleasel¹ | Stuart Lane¹  |
Annette Burgess¹  | Kellie Charles^{1,2} | Rosa Howard¹ | Deborah O'Mara¹  |
Inam Haq³ | Timothy Rutzou⁴

¹Faculty of Medicine and Health, Sydney Medical School, Education Office, The University of Sydney, Sydney, New South Wales

²Faculty of Medicine and Health, Sydney Pharmacy School, Discipline of Pharmacology, The University of Sydney, Sydney, New South Wales, Australia

³Faculty of Medicine and Health, The University of Sydney, Sydney, New South Wales, Australia

⁴School of Medicine, The University of Notre Dame, Chippendale, New South Wales, Australia

Correspondence

Priya Khanna, Faculty of Medicine and Health, Sydney Medical School, Education Office, The University of Sydney, Sydney, NSW 2006, Australia.

Email: priya.khanna@sydney.edu.au

Abstract

Background: Fundamental challenges exist in researching complex changes of assessment practice from traditional objective-focused ‘assessments of learning’ towards programmatic ‘assessment for learning’. The latter emphasise both the subjective and social in collective judgements of student progress. Our context was a purposively designed programmatic assessment system implemented in the first year of a new graduate entry curriculum. We applied critical realist perspectives to unpack the underlying causes (mechanisms) that explained student experiences of programmatic assessment, to optimise assessment practice for future iterations.

Methods: Data came from 14 in-depth focus groups ($N = 112/261$ students). We applied a critical realist lens drawn from Bhasker's three domains of reality (the actual, empirical and real) and Archer's concept of structure and agency to understand the student experience of programmatic assessment. Analysis involved induction (pattern identification), abduction (theoretical interpretation) and retroduction (causal explanation).

Results: As a complex educational and social change, the assessment structures and culture systems within programmatic assessment provided conditions (constraints and enablements) and conditioning (acceptance or rejection of new ‘non-traditional’ assessment processes) for the actions of agents (students) to exercise their learning choices. The emergent underlying mechanism that most influenced students' experience of programmatic assessment was one of balancing the complex relationships between learner agency, assessment structures and the cultural system.

Conclusions: Our study adds to debates on programmatic assessment by emphasising how the achievement of balance between learner agency, structure and culture suggests strategies to underpin sustained changes (elaboration) in assessment practice. These include; faculty and student learning development to promote collective reflexivity and agency, optimising assessment structures by enhancing integration of theory with practice, and changing learning culture by both enhancing

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Medical Education* published by Association for the Study of Medical Education and John Wiley & Sons Ltd.

existing and developing new social structures between faculty and the student body to gain acceptance and trust related to the new norms, beliefs and behaviours in assessing *for* and *of* learning.

1 | INTRODUCTION

Traditional assessment systems that emphasise the assessment of learning through summative high-stakes decision making have been critiqued for providing insufficient information about the complex competencies medical graduates require for entering rapidly changing health systems.¹⁻⁶ As an alternative, programmatic assessment provides an information rich, timely and sustainable process for strengthening the attainment and assessment of programme-level learning outcomes. It provides design principles around three key functions of assessment: promoting learning (assessment *for* learning), enhancing decision making about student progression (assessment *of* learning) and quality assuring the linkage between curriculum and assessment.⁷ Programmatic assessment supports assessment *for* learning by using purposefully selected multiple assessments combined over a period of time to create a longitudinal flow of triangulated information about a learner's progress in various competency outcome areas.⁸ Collecting and collating these data points not only provide a basis for collective decision making on student progress by faculty (assessment *of* learning) but provides a rich source of individualised feedback to learners (assessment *for* learning).⁷ The underlying theory and principles of programmatic assessment have been described in detail in the literature.⁷⁻¹¹ Notwithstanding, there is limited understanding of implementation approaches taken across different contexts that involve complex, dynamic and multilevel systems.^{12,13} Few studies provide empirical data supported by theoretically informed explanations of how programmatic assessment is working, for whom and in what context?¹⁴⁻¹⁶ This pragmatic approach is used in critical realist and realist evaluation,¹⁷ contrasting with traditional approaches to assessment research that typically asks, "what works?"¹⁸ There is thus a need to further develop appropriate research methodologies to ensure researchers are asking the appropriate questions when considering the impact of a complex educational intervention such as programmatic assessment.¹⁹ Without empirical data, it is difficult for educators to make informed decisions about introducing programmatic assessment, where the prevailing experience of assessment is often traditionally based.

In this paper, we extend current research on programmatic assessment by exploring the notion that findings from a critical realist (CR, hence forth) theoretical framework can provide insights into how design and implementation issues related to programmatic assessment can be optimised for future iterations. An opportunity to study this arose when a purposively designed programmatic assessment *for* and *of* learning was implemented in the first year of a new graduate medical curriculum at a research-intensive university in Australia. We wished to explore which elements of programmatic assessment seemed to be valuable for students' learning, under what

circumstances, and why this was so. To explain our CR approach for meeting our research goals in this context, we set out our overall programmatic assessment design emphasising the initial theories of how it was intended to work in our research context. Then, we describe our theoretically driven methodological approach and set out our overarching study aims and research questions.

1.1 | Research context

A new 4-year graduate-entry MD curriculum commenced in 2020 for 261 Year 1 students. It involved several changes from the prior curriculum including enhanced and diverse clinical immersion, a flipped classroom approach to content delivery, and horizontal and vertical linkages of curricular themes. Introducing programmatic assessment was a complex intervention involving a significant shift from the previous system of assessment that was traditional in the sense of having several formative assessments and major summative assessments such as written tests and the objective structured assessments of clinical skills. The programmatic system was devised through a series of local workshops, consultations with leading assessment experts in the Netherlands, Australia and New Zealand, and was cognisant of the relevant literature. However, cohort size, faculty experience with previous summative assessment frameworks, and local university assessment regulations and requirements required several contextualised and pragmatic adaptations to the theoretical principles. The COVID-19 pandemic influenced the implementation of several aspects of the new curriculum in terms of a shift towards online teaching-learning modes but did not significantly impact the programme theories underpinning both the curriculum design and programmatic assessment.

1.2 | Initial programme theories of implemented programmatic assessment

Our version of programmatic assessment was designed to align with complexity-consistent views of clinical competence. In considering a systems approach to overall programme design, a programmatic assessment approach provided one of several integral components that made up the features of the curriculum.²⁰ The programmatic assessment was intended to strengthen both the learning and decision-making functions of the prior assessment system.^{8,9} It included various new and revised assessment tools and improved structure with clear rules for submitting completed assessments and expectations of student behaviours. The argument for the validity of our programme of assessments was based on the carefully tailored

combination of various assessment instruments depending on the specific purposes within the overall programme.¹¹ Fairness was addressed from a perspective of equity, that is, all learners receiving the same quality of assessment.²¹ Information about learners was collected (longitudinally) and collated (triangulation) within a student progress record (SPR) that constituted a bespoke ePortfolio. This consisted of three broad elements: students' understanding of basic and clinical science knowledge, competence in clinical skills, and professionalism related aspects. The details of the key elements of the programme of assessment and their relation to the initial programme theory are given in Table 1.

1.3 | CR research framework

In viewing programmatic assessment as a complex social phenomenon, we used a CR stance to unpack and understand the complex relationships and causal mechanisms (ways of working) between the underlying design, the contextual implementation of programmatic assessment and their impacts on student learning. To the best of our

knowledge, CR has not been applied empirically to consider assessment systems within medical and health science education. CR might be a relatively new paradigm for many health professional educators more familiar with traditional positivist and interpretivist positions. Each of the three paradigms has a distinct position as to how the reality of any research phenomenon is determined.^{35,36} Positivism and social constructivism assume reality to be 'flat' and reduced to human interpretation and thus offering limited perspectives of the research phenomena. CR, on the other hand, assumes reality to be stratified and causally efficacious (an ability to cause an effect or outcome) and can be understood through a broader range of inferential techniques than induction.³⁵

Our research framework in this study was shaped by our previous work,¹⁸ which used a CR perspective to address programmatic assessment derived from two perspectives: first, Bhaskar's stratification of reality into three domains (the empirical, the actual and the real) (see Figure 1); and second, Archer's theory of structure and agency.³⁷ Bhaskar's concept of stratification allowed us to disentangle three intersecting domains of reality that shape the student experience of programmatic assessment: the empirical (data gathered from

TABLE 1 Features of the implemented programmatic assessment and their relation to initial programme theories in Year 1 of a graduate entry programme with the purpose of strengthening the learning and decision-making functions of assessment

1. Assessment formats

- a. Understanding and application of basic and clinical sciences knowledge was assessed using written tests administered four times a year.²² We anticipated that such a progressive approach would allow rich feedback for students, promote deep rather than superficial learning and provide timely monitoring for applied knowledge on all content areas of the curriculum.²³
- b. In addition to assessing clinical competence using objective structured clinical examination (OSCE) style clinical assessments, work-based assessments captured the performance of students over a period of time. We anticipated that these formats would be appropriate to detect issues in the traditionally difficult areas to assess, such as communication skills and professionalism.²⁴
- c. In the assessment of professionalism, assessors provided judgements on individual student's written reflections aimed at promoting self-regulatory behaviour. This also included assessment tasks on curricular themes such as indigenous health and interprofessional learning activities developed and assessed longitudinally. Additionally, a student progress record documented any late submissions. These assessments were guided by Hodges et al.'s²⁵ three levels of assessment or professionalism: individual, the inter-personal and the societal-institutional.

2. Learning advisor system

- A key element of supporting the programmatic assessment and encouraging students to be self-monitoring and fostering other elements of professionalism was the learning adviser system where a group of five to six students were allocated to a clinician from a teaching hospital.^{26,27} The discussion was focused on students' learning plans and progress records.

3. Proportionality

- In adapting the principle of proportionality,^{28,29} the stakes of decisions about a student's progress were intended to be proportional to the credibility or richness of information.²⁹ Students and assessors were informed that individual assessments in programmatic assessment were perceived as low or medium stakes and were intended to optimise self-directed learning.

4. Decision making and progression

- Decision making about student progression was managed by a portfolio advisory group meeting as appropriate to consider mid-year progress, the need for remediation, and end of year meetings to determine progression to the next year. The longitudinal nature and the information richness of the triangulation process were intended to make programmatic assessment defensible for high-stakes assessment of learning.³⁰

5. Remediation

- Given the judgement on student progression are akin to 'diagnostic' decisions, a remediation process was designed to support students who performed below the expected level. The decision of sufficient competence to go to the next stage of training was thought to be prognostic.¹⁰

6. Integration with other aspects of the curriculum

- Key curricular elements provided important assessment tasks 'as' and 'for' learning. Team-based learning (TBL) was the dominant teaching activity in Year 1 providing a collaborative approach wherein students worked in small teams on authentic cases. The sequencing of activities encouraged students to apply conceptual knowledge through a series of steps involving preparation, readiness assurance testing, feedback and the application of knowledge through clinical problem solving activities.^{31,32} It was supported by the flipped classroom method of learning and teaching.³¹⁻³⁴ Thus, readiness assurance tests and explanatory mechanistic diagrams of the specific clinical problem in the TBL were included in the student progress record.

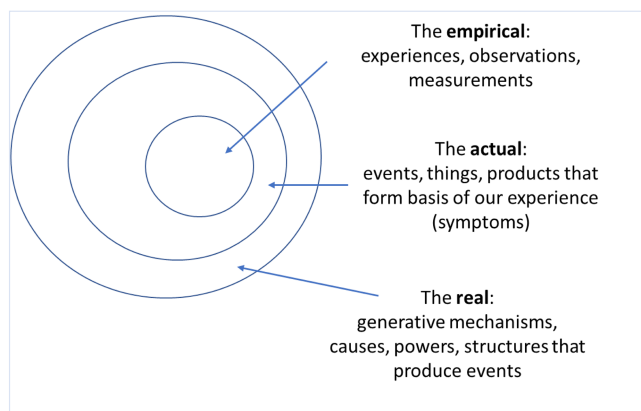


FIGURE 1 Bhaskar's critical realist lens on domains of reality [Color figure can be viewed at wileyonlinelibrary.com]

observations and experiences), the actual (events or non-events that students report within the assessment programme) and the real (underlying causal structures and mechanisms).^{36,38,39} See Figure 1.

The analogy of clinical diagnosis can be used to illustrate how reality can be stratified into three domains (the empirical, the actual and the real). The experiences associated with programmatic assessment (i.e., actual events recorded at the individual level) are akin to the symptoms and signs that a patient might present to a doctor.⁴⁰ The empirical level captures the experiences of the person (patient) that are akin to a history and examination in providing measurable and assessable data. Causal structures and mechanisms are real, distinct and potentially different from both the actual and the empirical and are akin to the underlying pathology and diagnosis of the patient.

Archer's CR perspective allowed us to explore the complex causal interplays between structure, culture and agency that might contribute to and impact the transformation of assessment practices.³⁷ In our research context, structure describes the policies, positions, resources and practices, whilst culture describes the system of meanings, beliefs, norms and ideas associated with systems of assessment. The complex interplay over time between structure and culture, together with human agency,^{41–43} inevitably results in cyclical dynamic change referred to by Archer as 'morphogenesis', or staying the same, referred to as 'morphostasis'⁴⁴ (see Figure 2).

1.4 | Study aims and research questions

The purpose of this study was to explore from a CR perspective, students' perceptions of which elements of programmatic assessment influenced their learning and why.

Our specific research questions were as follows:

1. To what extent did the features of the new assessment system influence students' ability to direct their learning needs?
2. What were students' experiences in navigating assessment formats, rules and practices?

3. How did the interactions with various entities within the new curriculum such as faculty and peers influence students' engagement with the programme?
4. What were the underlying explanations of students' perceptions and experiences and how might they influence growth and sustainability of programmatic assessment?

These questions are important as they provide rich and theory-based explanations of what is really working, how and why it is working, to optimise the programme of assessment and the student experiences in further iterations.

2 | METHODS

2.1 | Study design

We addressed our research questions using a qualitative methodology drawing on critical realism to explore the influences on students' perceptions of various aspects of the programmatic assessment.

2.2 | Data collection

Data were collected from 10 (labelled A–J) in-depth focus groups (total $n = 112/261$, 43% of student cohort) across six of the seven teaching hospitals in which students were based (in range 15–52 students per site) for 1 day a week during the first year of the programme. Cohort demographics are illustrated in the Table 2. For recruitment, students were made aware of the study and invited to attend a focus group at their home clinical school. The initial sampling strategy was modest and anticipated that around 20 students would have provided sufficient information power⁴⁵ to report on the learning advisor system. However, additional focus groups were arranged to account for student interest in having their voices heard about the assessment changes.

Having been consented, interviews lasted from 40 to 60 min each and were conducted by PK, CR, AB and SL. An initial interview guide was developed from insights from the literature, prior theorising¹⁸ and the authors' experience of an initial learning advisor pilot evaluation in the previous year. In the interview schedule, questions were designed to elicit students' experiences of programmatic assessment focussing on the ways in which learning advisors supported or challenged student learning *as* and *for* assessment. However, during focus groups, it became clear that students had significant insights they wished to share on not just the learning advisor system but with the programmatic assessment system as a whole. Accordingly, focus groups were conducted as a conversation allowing students to elaborate on their perceptions of the programmatic assessment system including judgements of student competence and the fit of assessment with learning activities. Audiotapes were deidentified, transcribed verbatim and stored on the university data protection facility.

FIGURE 2 Archer's theory of interplay between culture, structure and agency when viewing a social system [Color figure can be viewed at wileyonlinelibrary.com]

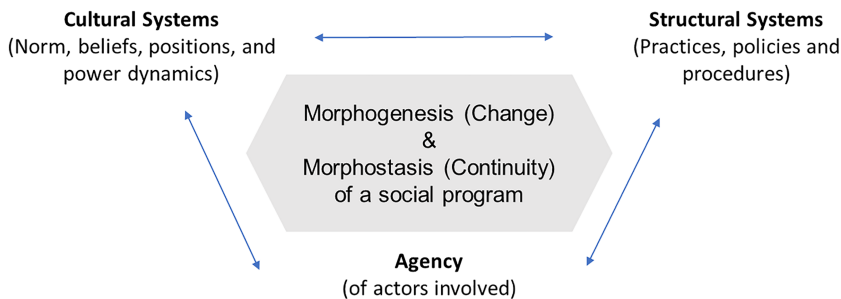


TABLE 2 Year 1 cohort demographics ($n = 261$) by gender, prior degree and application status

Demographic	Number of students in year	Percentage of cohort
Gender		
Male	148	56.7
Female	113	43.3
Prior degree		
Science	192	73.5
Arts/Business/Law/Maths	31	11.9
Health Professions Education	26	10
Not stated	12	4.6
Application status		
Local	208	79.7
International	53	20.3

2.3 | Data analysis

We wished to unpack the underlying causal mechanisms that shaped the student experience of programmatic assessment. In line with the CR framework, data analysis and synthesis were dynamic rather than linear and involved three phases of inference undertaken iteratively. To ensure understandings applied to the full data, differences in researcher perspectives were negotiated through meetings (face to face, video conference and email) and using whiteboarding of the research materials. Data were managed using the qualitative data analysis program NVivo (Version 12), (QSR International Pty Ltd. 2020).

2.3.1 | Phase 1: Induction

The initial focus of our iterative inductive analysis was around unpacking and describing the cohorts' experience of programmatic assessment and its related curricular elements, without being tied to a specific theory.³⁵ We (CR, PK and TR) reviewed the raw data to iteratively identify general and emergent patterns, connections, similarities and variances. At this point, we noticed the internal limitations of induction in understanding the often conflicting causes that appeared to underlie the students learning and assessment experiences.³⁵

2.3.2 | Phase 2: Abduction

Abduction provided a means of forming associations in the data that went beyond initial pattern recognition to give a more comprehensive understanding of the emergent patterns.^{35,46} To illustrate our method, abduction linked concepts that might have been understood within a particular context (e.g., team-based learning [TBL] as a teaching activity) to the new context of programmatic assessment (TBL as an assessment *for* learning) through redescription or recontextualization. The initial inductive coding of the data was re-examined and recoded through abduction and then reorganised and recontextualised into a CR-based conceptual map of stratified domains of reality as actual, empirical and potential 'real mechanisms'. Data coded as 'actual' (akin to clinical symptoms)⁴⁰ reported by the students whether observed or not. This included the way students perceived everyday assessment activities, their achievements and how it made them feel. The coding for the 'empirical' (akin to clinical examination and investigations) included students' observations, perceptions and reflections of various aspects of the new assessment system. At this stage, the 'real' (akin to a differential diagnosis) could only be coded as potential mechanisms that explain why and how the actual and empirical came to be, but not which were the key mechanisms.

2.3.3 | Phase 3: Retroduction

Retroduction, a key component of CR methodology, involves causal explanation of the data to unpack the basic conditions, structures and mechanisms that cannot be explained at the actual or empirical levels alone.³⁵ We re-coded both the converging and conflicting potential 'real' mechanisms in the light of Archer's theory of the interplay between culture, structure and agency.^{37,41} (Figure 2) Using the three modes of inferences, the evolving explanation of findings moved between the actual, the empirical and the real, taking account of the context of our implementation of programmatic assessment and the underlying programme theories.^{46,47}

2.4 | Team reflexivity

Our team comprised multidisciplinary and experienced researchers, clinicians and clinical scientists, who were collectively and directly

involved in creating and implementing the programmatic assessment. A social scientist, not involved in the programme, helped construct meanings in this research⁴⁸ by providing differing insights into the data. Some of the authors were familiar with critical realism prior to the study. Reflexivity was promoted through meetings and via email sharing our internal conversations, the reflexive deliberations through which the individuals address and prioritise their concerns about what the data said about changing assessment practice.³⁷

3 | RESULTS

We present our findings in two parts: First, an account of the ‘empirical and actual’ using induction and abduction that addresses the first three research questions; second, an account of the unpacking of the ‘reality’ of programmatic assessment for the students using abduction and retrodution that addresses our fourth research question.

3.1 | Part 1: The ‘empirical and actual’

The empirical and actual levels of reality included students’ personal experience and interpretations of actual events they experienced in relation to first, their curricular components such as learning and teaching activities, or various assessments; and second, their perceptions of the culture of the learning environment, as reflected in their recall of communication, beliefs and norms in regard to programme requirements. Considering our research questions, we developed three themes in relation to student experiences of programmatic assessment: (1) enacting learning choices (agency), (2) navigating the assessment system (structure) and (3) building a cultural system.

RQ 1: To what extent did the features of the new assessment system influence students’ ability to direct their learning needs?

3.1.1 | Enacting learning choices (agency)

This theme describes the factors that mediate the interplay between structure (the assessment rules, practices and resources that may enable or constrain action) and agency from the perspective of how and what choices students have in their learning. By ‘choices’, we mean the perceived degree of freedom that allowed students to identify their learning needs and direct their learning accordingly.

Students reported various influences within the medical programme, and the programmatic assessment specifically, that shaped the degree to which the structures in place gave them choice in what and how they learned and by when. For example, student agency within the learning adviser system was expressed through having a

long-standing relation with teaching faculty, in which a professional conversation was a constructive influence on their professional development journey from being a student and becoming a doctor. For most students, the learning advisor system supported their own self-efficacy and self-regulation through analysing their strengths and weakness in their learning and devising a personal learning plan to work on those weaknesses.

I definitely liked having the idea of having the (development) plan going in, because it sets up what you want to talk about. And then I also liked having the actions at the end, and then having the update to the plan based on what you discussed in the meeting, because it really forced you to set actions for yourself after the meeting. (C2)

However, for some students, the learning advisor component, in its current form, was perceived as another assessment to complete as part of programme requirements. Thus, imposing an additional meeting to “catch-up, check in, ask a few questions” (C) to evidence their engagement with the learning advisor process, rather than supporting their learning. At times, students perceived that certain structures of the assessment system worked against their learning rather than supporting it, such as a perception that mandatory attendance rules were a marker of professionalism.

I find it frustrating not that they are forced to meet attendance, it's like, for example, at the start of the year, they came in and randomly did a manual check of who was in there because they did not trust that people were doing the QR (Quick Response) code properly or even in our Zooms (video conferencing software) at the start. (H)

Students talked about elements of the assessment system, which they thought were impacted by resource issues, for example, the process by which completed assessments were uploaded in the learning management system and the delays they perceived in getting feedback on individual and team assessments.

The reasons that they are not meeting these deadlines is probably because they are overworked, and understaffed, and starting a new course you need more support, not less. Which is kind of through the grapevine what we have heard has happened. And it has affected our learning. It has impacted on us, and it has definitely changed the quality of the education we are getting (J)

Notwithstanding, the programmatic assessment system was recognised as having the potential to develop over time, for example, developing the notion of a learning advisor being a mentor on the student journey through the medical programme.

I use my learning advisor not just for the curriculum itself, but also as a bit of career coach. So, I think there's a lot of value to be had there, and I did that because I come from a different profession previously, and my firm was quite keen on having a career coach that will suit – mentor you on the way, and connect you with the right resources, if you ask for it. But that's only possible when the students feel empowered to reach out for that kind of help. (A)

In summary, students made most sense of their agency related to three factors: first, their own motivations and self-regulation from having a study practice that they felt helped them in progressing through the course; second, with professional conversations with clinicians as learning advisors; third, from constructive and individual feedback around individual or team assessments. Learner agency, to some extent, was constrained by navigating the IT systems, negative communication experiences with faculty and a perceived lack of resources to deliver the intended programme.

RQ 2: What were students' experiences in navigating assessment formats, rules and practices?

3.1.2 | Navigating assessment structures

This theme describes the dynamic interplay between student expectations of the programmatic assessment, the faculty's intended implementation and the institutional delivery of the programme. Critical structures of programme functioning were found in their comments about the judgement and decision-making process on the completed collection of assessment tasks within the student progress report, especially the remediation process, and the professionalism-related issues. In expressing their opinions, students showed fallible misunderstandings about the programmatic assessment, shaped by their prior conditioning to traditional assessment systems.

There were a number of tensions in both learners and faculty around the narratives about programmatic assessment, for example, what was understood by the terms 'formative' and 'summative'; interpretation of 'stakes'; and assessment *for* and *of* learning.

I get it; the difference between formative and summative, and that everything is formative in this course until it is summative. Problem is they do not say when that is so you could be sweating on a minor thing to know if you have passed the whole year. (G)

The stakes or weightage describes the degree to which faculty decisions about student's progress are proportional to the credibility of information. There was much uncertainty amongst learners' incomplete and often fallible perceptions of assessment as to what the stakes of a particular assessment were compared with other assessments included in the SPR.

I would say at the beginning of the year they did say everything is—they said something along the lines of everything is of equal weight, or nothing is weighted more than the other. So, I guess in that line, they are trying to get the—which I think they have at least held true to, they have not let us know anything is weighted differently. But if that's true or not, I do not know. Behind the scenes, is the [written test] weighted more than an Anatomy Spot Test? (C)

Of the individual elements of the programmatic assessment, most students felt the continuous testing in the written assessments seemed to "have been a fair way to assess us" (J), as well as "less stressful than having a barrier" (F). The progressive assessments provided an indication of where they were in terms of progress in the first year of the programme and seemed to be supportive to their learning.

In the work-based assessment, COVID had impacted the affordances of the workplace for observation and feedback, but overall was regarded as working as intended and giving useful feedback on student progress in developing clinical skills, whilst providing immediate feedback.

The marking scheme was really good in that it's quite generalised, so it's not very specific. So, I feel like they did say for this year it's more about building confidence by doing the history and physical exams, and I think that those really built up the confidence. I also really liked how in our first block in respiratory, for the physical exam it was quite small. It was kind of just doing peripheries, but then they just built up to something like maybe the whole abdominal exam. (C)

The collaborative production of mechanistic diagrams in the TBL sessions was seen as a useful learning endpoint of the TBL process and a good match to programme theory. However, when recrafted as a team-based assessment, as part of COVID adaptations, the students perceived them as a high-stakes assessment, strongly detracting from overall learning. Students explained this was mainly because of an uneven balance between the time invested and their learning gains, a matter of extraneous cognitive load.⁴⁹ For some groups, collaborative teamwork became focussed on achieving higher marks at the expense of the overall learning value. This was on a background of student uncertainty in how to interpret the TBL-related marks and what the standard expected might be.

When we started doing the mechanistic diagrams, I feel like—when we are doing it in class, they are very simple. And then as soon as they had to be handed in and marked, we were expected to have a high level of complexity and large amounts of detail within this diagram, which of course adds more time, and you have to spend more time thinking about it. (B)

Similarly, the individual readiness assurance tests (iRATs) as part of the TBL sessions were seen as a useful motivator of learning. However, as indicators of progress in learning, iRATs were considered to have too much perceived importance, given they were intended to indicate readiness to learn before the TBL rather than a measure of satisfactory achievement of learning.

The collective faculty view on assessing professionalism including recording late submission of tasks in the SPR was problematic for students. They worried the process felt punitive as they feared being judged in breach of professionalism for a minor issue when submitting one of multiple assessments. Further, trivial breaches could remain in the student record and bias faculty impressions of them. Students appeared more comfortable with viewing unprofessionalism as a lapse to be worked on,²⁵ with additional support and work, rather than being viewed as a professional breach.

With assessing professionalism, I feel like it's sort of more something to be lost, rather than having to assess someone as being professional. But then like sometimes things might happen throughout the year, the way someone acts in lectures, or at clinical school, like, there's instances where I feel like they are displaying unprofessionalism, and that is sort of what should be reflected instead. (A)

The implementation of the remediation process led to uncertainty. This was amplified by the claimed lack of knowledge amongst the students about the decision-making and progression process and the role of the portfolio advisory group in the end of year review of the SPR.

It's hard to know, not going through it (remediation), it's hard to know how much was disclosed to the people that did, but it has not seemed all that transparent in terms of you do not understand, if you were to fail an exam, what the consequences are and I was under the impression to begin with that it was quite a supportive process, and that you'd be given the resources and told how to improve so that then you could, with the aim of completing the year. But it seems like that's not reality necessarily. (E)

In summary, in terms of programme theories around the collection and collation and the reporting of assessment data, the progressive testing of basic and clinical science and the work-based assessment had worked largely as the programme theories had predicted they would. The difference between the affordances of TBL as a key learning method and its use in a different context, as an assessment, had a significant proportion of students challenging the value of including, for example, readiness assurance testing in the SPR. The assessment of professionalism was perceived as simplistic and limited to the lateness of assignment submission. Problems in socialising the way in which decision-making and progression rules

operated were manifested in students' uncertainty of the stakes of differing assessment formats and the utility of the remediation system. That was stressful for some students.

RQ 3: How did the interactions with various entities within the new curriculum such as faculty and peers influence students' engagement with the programme?

3.1.3 | Building a cultural system

This theme describes the system of culture prevailing within the programme, the sense of students and faculty attempting 'to learn how to learn together'.⁵⁰ Programmatic assessment was a new experience to both students and faculty. The theme also includes the contribution of the emotional turmoil of the students, the notion of unproductive work and the student perceptions of 'being a difficult cohort' in the learning contexts they found themselves in. Their sense of the learning cultural system impacted some students' assessment as learning in a number of ways.

And even—yeah, it's just hard to stay motivated when you feel like people aren't on your side, when they are supposed to be on your side. So, it sounds a bit whingy, I know, but I just think it's overall, like, brought down the mood of this cohort. And I do not think I'm exaggerating when I say that. (J)

For some students, there was a perception of a negative learning culture or even a 'blame culture' (J) where the ethos of the programme was perceived as one of surveillance, more appropriate to undergraduate students straight from school. Students recognised that their fellow students, perhaps excluding themselves, were being resistant to change in the context of learning medicine.

But then they also want to punish you punitively in assessments as if you are an under-grad. Now, if you are a 17- and 18-year-old, you need to be kind of whipped into shape to learn how to exist in a tertiary education system. Fine. I'm actually okay with first year subjects being a bit punitive. (B)

There were a few ways that the student found as workarounds to get the information they felt they required from faculty. One of these ways was through the student representatives, for example, "relying on student reps (year student representatives) to post up on the Facebook (private Year 1 social media) group" (C). Others noted that the faculty communications were problematic.

I think student reps are doing an amazing job. But you guys should not be responsible for telling us what's going to be on our assessment. That should be on the faculty (J)

Students had experienced a range of differing situational factors impacting their participation in assessment and learning through their emotional turmoil.⁵¹

Well, for me personally, I went into the first, for example, the anatomy spot test, being really, really stressed. Spent a lot of time dedicating to studying for it, and kind of sacrificed a lot of my learning for the other content that was going to be assessed in the (written assessment). And then I found out after that that the reason that it wasn't a proctored exam is because it does not weigh enough. (J)

However, several students noted that a proportion of their peers were being unprofessionally 'rude', hiding behind their anonymity, and not taking responsibility for their feedback to the faculty. The solution appeared to be better communications between faculty and students and setting clear expectations, through multiple means of communication.

Just setting those expectations well at the beginning of the year. So, in your—in the foundation sessions, setting the expectations around professionalism, around participation, around what learning resources people should be engaging with, and assessing expectations around what students need to be doing supplementary to that, solves a lot of these problems. (A)

In summary, our data suggested that most students embraced the requirements for new ways of thinking about learning and its relationship with assessment. However, some students did experience what Durkheim called *anomie*, a breaking down of social norms (normlessness) regulating individual conduct in some of the communications with the faculty.⁵² It was made manifest in the talk of students, first with each other, and their student representatives. Second, through their communications with faculty via email, announcements through the learning management system, and clarification of information contained in student handbooks. In the students' social interactions about the programme with faculty, the perceived problems of the learning culture may have arisen because of the gap between the cultural goals espoused by faculty and the institutional means to deliver them.

3.2 | Part 2: Unpacking the real

RQ 4: What were the underlying explanations of students' perceptions and experiences, and how might they influence growth and sustainability of programmatic assessment?

Explanation of how, why and what features of the programmatic assessment worked for the students from a CR stance involved unpacking the underlying mechanisms behind students' perceptions

and experiences of how the assessment system impacted their learning.^{36,53} Archer's theory of morphogenesis (introduced earlier)^{44,54} provided us with deeper insights into disentangling and understanding causal linkages (mechanisms) underpinning the three-way interplay between structure, culture and agency in the context of implementing programmatic assessment. The morphogenetic approach allows for a stratified account of structures, culture and agents, as each has emergent and irreducible properties and powers (i.e., causal mechanisms) that explain the student experiences of programmatic assessment.⁵⁵ These mechanisms only manifest themselves under specific conditions (i.e., constraints and affordances) when changing to programmatic assessment to further elaborate (morphogenesis) or to resist the change (what Archer calls morphostasis⁵⁴). The cyclical and ongoing process of morphogenesis in programmatic assessment depends on maintaining optimal tensions between 'conditioning' (consensus and integration of new practices with previously held beliefs) and 'elaborations' (acceptance and adaptability of the new features).

In reintroducing the 'real' from our findings at the empirical and actual level, the structure and culture systems associated with the programmatic assessment have distinct but related causal powers (see Figure 3). In particular, retroduction highlighted first, the centrality of reflexive deliberation of the student cohort in maintaining a truly programmatic agenda; second, the importance of attending to (student-led) changes in the local structures and cultural systems surrounding assessment to facilitate the morphogenesis (rather than morphostasis) required for the sustainability of programmatic assessment.

An important example of balancing the interplay between assessment structures, the cultural system and student agency, was the valuable professional conversations with learning advisors concerning broader aspects of becoming a clinician. This functioned to enhance the agency of most students in their education. At the same time, the students were also conditioned to sustain and further elaborate the intended purpose of programmatic assessments, by calling for a more mentorship type relationship rather than a purely assessment focussed one that would only facilitate their understanding of medical knowledge.

In contrast, getting the balance uneven potentially leads to situations of morphostasis where the implemented programme of assessments remains problematic due to the loss of student agency and the return of the traditional assessment structures and culture. This can lead to a sense of 'anomie'⁵² for many of the students, that is, a normless state leading to dysfunctionality and inactivity in the intended purpose and actual delivery. One of the consequences is threats to the inherent morphogenesis of programmatic assessment due to a mismatch between assessment ideals and practices.

Examples of uneven balance include the mismatch between the intended theory and the implemented version of assessment components, impacting both the structure and culture of assessment. Student perceptions about the supportive versus punitive purpose of remediation, for example, served as potential constraints to students' agency. Similarly, punitive perceptions of the assessment of professionalism, where the ePortfolio was experienced by students as a tool for surveillance and control, diminished rather than enhanced

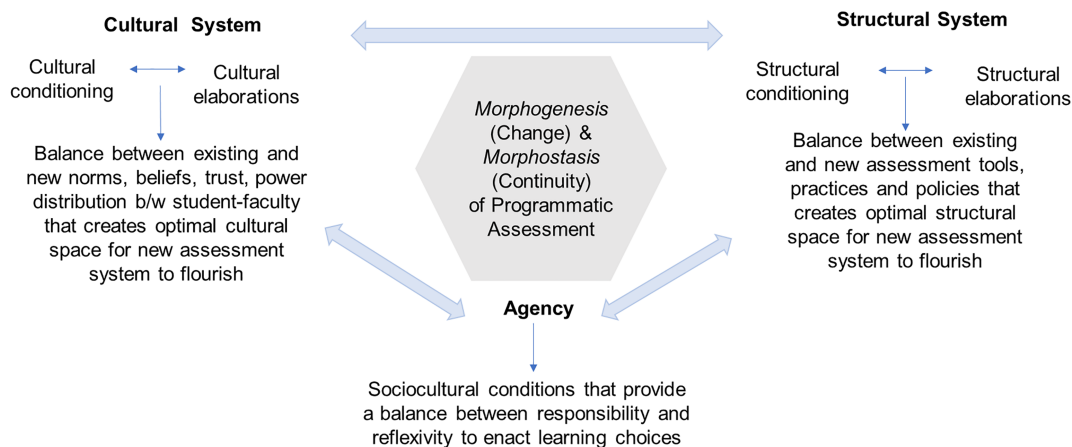


FIGURE 3 Balances in the complex intersection between structure of assessment, the cultural system and agency allowing the necessary conditions for programmatic assessment to flourish [Color figure can be viewed at wileyonlinelibrary.com]

reflexivity and agency by creating a sense of passivity and a feeling of powerless and passive inactivity (anomie).

In summary, providing balance between structure, culture and agency to sustain morphogenesis in developing and elaborating programmatic assessment relies on optimal distribution of power dynamics and trust between faculty and the student body. This emerged as central to implementing successful changes with a lack of balance leading to the counterproductive sense of a hidden curriculum within the programme of assessments.

4 | DISCUSSION

4.1 | Summary of key findings

We sought to understand how elements of a newly introduced programmatic assessment worked for the students in terms of enhancing their learning under what circumstances and why. We used a CR stance based on three domains of reality (empirical, actual and real) allied to a cyclical dynamic model of change (morphogenesis) to unpack the complex interplay between the structures of assessment practice, the conditions of the culture system and learner agency. Our key finding was discovering underlying mechanisms that were both explanatory of student experiences and suggestive of ways in which future iterations of the programme of assessments could be optimised. The model that emerged from this study is given in Figure 3, which is a way of visualising the causal mechanism as one of balance between structure, culture, and agency. An example of a constructive balance was found in the operating of the Learning Advisor system. A lack of balance led to the sense of a hidden curriculum for students within the programme of assessments.

Considering programmatic assessment as a complex social change, the associated structures and culture provided conditions (in the form of constraints and enablements) and conditioning (in the form of acceptance of new 'non-traditional' assessment processes) for the actions of agents (student and faculty choices). These

interactions resulted in traditional assessment practices being transformed (elaboration) towards programmatic assessment. However, there was also indication of how the assessment practice could remain traditional. The process of morphogenesis and morphostasis provided an understanding of the cyclic nature of changes within the programmatic assessment that are sustained, continued, and yet which can easily revert back to the original features.⁵⁴ A mismatch between intended programme theories underlying various elements of programmatic assessment and the students (often) fallible perceptions in their experience of the programmatic assessment can be a threat to sustaining new assessment practices and optimising student learning and reflexivity. The dynamic flexibility of the CR approach proved helpful in understanding how the various elements of programmatic assessment (such as assessment formats, learning advisor system, remediation and progression decision making) were interrelated, such that change to one area impacted change in another.

4.2 | Comparison with existing theory and literature

Our findings add to the existing debates about programmatic assessment in health professions education by suggesting that there are important and often neglected causal mechanisms that impact the students varied experiences during the implementation of programmatic assessment. This adds to the existing literature on implementation approaches involving complex, dynamic and multilevel systems.^{12,13} Our findings add to the overarching causal mechanisms that can explain student experiences of programmatic assessment.^{12,13,56} Our research also extends previous theoretical work in using a CR stance.¹⁸ It does so by providing empirical data and identifying potential mechanisms that are explanatory by revealing the delicate balances between students' agency and the rules and regulations of assessment and the local cultural system. It extends current thinking on culture and cultural change in health professional education^{57,58} by identifying a culture system that is linked with structures of assessment

and student agency.⁵⁴ This research extends current theoretical thinking on insights gained from student agency.⁵⁹ In developing the comparison and extending theory, we discuss three areas that take account of the causal mechanisms and the dynamic interplay between structure, culture and agency in Figure 3.

1. Promoting agency through collective reflexivity.
2. Integrated and flexible approaches to assessment structures.
3. Addressing socio-cultural conditioning.

4.3 | Promote agency through collective reflexivity

Reflexivity is one of the central concepts in Archer's framework of CR. The interplay between people's 'concerns' (the importance of what they care about) and their 'context' (the continuity or discontinuity of their social environment) shapes their mode of 'reflexivity'.^{37,55} In our context, some of the programmatic assessment practices were firmly grounded in the theory of assessment of learning. Accordingly, opportunities for learners to be agentic⁶⁰ were somewhat constricted. A two-way dialogue between faculty and students might promote learners' agency including their ability to make choices during their experience of assessment and learning, facilitating assessment *for* and *as* learning.^{61–63} Others have noted that agency may be hindered by learners' perceptions of various data points in the assessment as high stakes and suggested that faculty can promote learner agency in safe and trusting assessment relationships.⁶⁴ Our data suggest that to promote learner agency, some of the assessment structures and the ways they integrate need to be changed. Students have direct responsibility in this regard to exercise their agency and engage with assessment tasks they had a hand in designing to facilitate their learning. The orientation for students undergoing a major learning transition from the traditional to a radical programmatic assessment system is important, so that students' level of preparedness, their agency in terms of engagement and learning is optimised.⁶⁵ For any mechanism to promote student agency whilst simultaneously adapting assessment structures, there has to be further investment in developing certain causal powers and capacities.³⁶ For example, there would need to be empowerment of the student body by faculty to contribute to a change in the assessment structures and promote student reflexivity. The degree of empowerment would depend on the local context and culture of programme delivery.

4.4 | Integrate and flexible approaches to assessment structures

Many elements of the programmatic assessment worked as intended, such as multiple written tests and the work-based assessment, and were perceived as useful indicators of student progression. In our study, a mismatch between intended programme theories underlying various elements of programmatic assessment (such as remediation designed to support learning) and the students' perceptions of the assessment (such as remediation perceived as punitive) suggested a

threat to sustaining new assessment practices and risk returning to previous traditional practices.

Although others have noted that assessment tasks can fail to challenge learners' sense of agency by virtue of not being understood or too complex, we found that, rather than complexity being an issue, some assessment tasks (for example in the team based learning) were felt to represent too great a workload and had a negative impact on learning progress.⁶⁶ The overall extraneous cognitive load of assessment requirements needs to be balanced in favour of learning, particularly in ensuring the integration of the different elements of the assessment as well as other aspects of curriculum.^{1,49}

The issues with mismatch of programme theory around proportionality or stakes of the individual assessment formats resonate with findings elsewhere. Schut et al.⁶¹ and Heeneman et al.⁶⁷ reported a mismatch between programme designers' intentions underpinning low-stakes assessments and students' perceptions of these assessments as summative or high stakes. The moment 'stakes' are assigned to assessments; they are perceived as 'summative' by learners that may interfere with the intended educational effect and change the nature of student learning.^{61,67,68} In our view, the terms 'stakes' on individual assessment are best applied to the decision-making process on the whole of the SPR.¹⁸

Implementing change to the culture of assessment *for* and *as* learning requires open interactions between the assessment programme, the assessors and the students,^{9,10,69} creating an enhanced relationship between the intended design and assessment outcomes.^{7,20} Sufficient resources are likely to be needed for enabling structures to support learner agency such as learning advisors.⁶¹ There is an important role for Faculty development in socialising the decision-making and progression rules as most student and staff have a history of working with traditional approaches.

4.5 | Address sociocultural conditioning

According to Archer's notion of morphogenesis, the processes of change in a learning culture are sustained for both the students and the structures of the assessment programme in interlocking and temporally complex ways.⁴⁴ Students acquire new norms, new communities of peers and power relationships. Addressing the structural and cultural concerns of students in engaging with programmatic assessment will lead to elaboration and refinement of the design of assessments. Neglecting this may reset the culture back towards the norms of traditional assessment. In short, programmatic assessment requires not only structural changes to the curriculum but also a simultaneous cultural shift.

It will always be a faculty concern if a significant proportion of students experience a hidden or unintended curriculum, when undertaking assessment reform. There has been discussion of the hidden element of assessment in the literature.^{70,71} Our data suggest that the student 'anomie' with the programme was principally around the symptoms and signs of students experiencing an overbearing structure of assessment that was a misfit for their expectations of assessment *for* learning. This was exacerbated by a perceived culture of suboptimal faculty

communication, an unfair expectation of student representatives to negotiate areas of perceived conflict and a sense of feeling assessment was a practice of surveillance of students rather than trust based. Although a small minority were subversive in their interpretation of their experiences of the programmatic assessment, the majority spoke of potential culture changes towards more student choice and expression of agency, and less rigid assessment protocols, an issue of structure. Changing the culture of the learning environment is best managed through developing a strategy for active student engagement with curriculum change, covering peer teaching and support programmes, governance processes and ultimately assessment change.^{72,73}

4.6 | Methodological strengths and uncertainties

Our study has several strengths and limitations when making sense of the findings. As a strength, this is one of the first theoretically rigorous accounts of an implementation of programmatic assessment. It goes beyond reporting student experiences by aiming to provide a much deeper explanation of students' perceptions and how they might influence growth and sustainability of programmatic assessment as an educational and social change. The use of critical realism and abduction and retroduction as inferential tools^{35,46} is relatively new to the health professional education research. Coding interview data at ontological levels (actual, empirical and real) and drawing meaningful inferences that go beyond dichotomous inductive–deductive modes not only added to methodological rigour but also provided more meaningful explanations of the findings. Lastly, by collecting the perspectives and experiences of 112 students across 14 focus groups, we claim a sample with sufficient information power given the narrow aims of our study.⁴⁵

In terms of methodological challenges, first, there was some overlap between inductive, abductive and retroductive phases of analysis with the process being more iterative than linear. Similarly, in practice, although there was some overlap between researchers as to what was considered empirical and actual, there was much greater agreement on what constituted the real; second, reconceptualising the data in terms of structure, culture and agency⁴⁴ might have benefited from considering staff agency. Although the focus of this study was the students' account, perceptions of faculty, who have their own agency, need to be considered in providing a richer understanding of the influence of structure and culture of the programme on student agency. Finally, we acknowledge that our analysis and theorising was based on a specific context where programmatic assessment was introduced as part of a new curriculum. Nonetheless, we believe our findings to be adaptable to others in differing contexts seeking to understand the implementation of programmatic assessment with large cohorts of students.

4.7 | Implications for practice and research

A CR approach based on the relationship between structure, culture and agency^{37,44,54} and stratified domains of reality^{36,38,39} can provide a methodology that is meaningful and adaptable for researching

assessment practice in health professional education by highlighting neglected areas of concern. For educators, our structure culture agency framework (Figure 3) can provide a toolkit which promotes a simplification of the complex educational and social changes when (re) designing assessment and curriculum. A CR approach can provide credible analysis for determining what might work across multiple contexts, what works for whom, and most importantly explaining how it works in terms of fundamental mechanisms. In this paper, we have focussed on student experiences, but such investigations can be extended to understand faculty's perspectives. A CR approach may help to explain how various entities within the medical school or university, for example, the faculty leadership groups, and teaching faculty develop and exercise causal powers that influence the flourishing of programmatic assessment. Research programmes based on CR perspectives can also enrich understanding of the fundamental principles underlying the link between programmatic assessment and curriculum thereby leading to local quality improvement and better adaptations of the principles in other contexts.

By extending the repertoire of approaches to researching assessment reforms, evaluation of complex educational initiatives like programmatic assessment can be unpacked and understood from the more fundamental perspective that CR offers. In future work, using CR ensures that intended outcomes are customised and adapted to local needs and contexts. This would ensure pragmatic implementation and long-term sustainability of programmatic assessment.

4.8 | Conclusions

Our study adds to debates on programmatic assessment by emphasising how the achievement of balance between learner agency, structure and culture suggests strategies to underpin sustained changes (elaboration) in assessment practice. These include; faculty and student learning development to promote collective reflexivity and agency, optimising assessment structures by enhancing integration of theory with practice, and changing learning culture by both enhancing existing and developing new social structures between faculty and the student body to gain acceptance and trust related to the new norms, beliefs and behaviours in assessing *for* and *of* learning.

ACKNOWLEDGEMENTS

We are hugely grateful to all the participants in this study for sharing their perspective. We are also grateful to the contributions of several members of staff associated with the medical school for their assistance with the study. Open access publishing facilitated by The University of Sydney, as part of the Wiley - The University of Sydney agreement via the Council of Australian University Librarians. [Correction added on 20 May 2022, after first online publication: CAUL funding statement has been added.]

CONFLICT OF INTEREST

All members of the authorship team except TR were involved in designing various elements of the programmatic assessment and so have a potential conflict of interest.

AUTHOR CONTRIBUTIONS

CR and PK designed the study and developed the critical realist methodology with TR and SL. PK secured ethics approval. PK, CR, AB and SL undertook the focus group interviews. All authors were involved in analysing and interpreting study data. CR and PK wrote the first draft, and all authors commented on and edited various iterations of the paper. All authors give their final approval for this version to be submitted.

ETHICS STATEMENT

This study was approved by the University's Human Research Ethics Committee (Approval No: 2019/328).

ORCID

Chris Roberts  <https://orcid.org/0000-0001-8613-682X>

Priya Khanna  <https://orcid.org/0000-0002-8441-2322>

Stuart Lane  <https://orcid.org/0000-0001-8650-5509>

Annette Burgess  <https://orcid.org/0000-0001-9617-3819>

Deborah O'Mara  <https://orcid.org/0000-0001-5131-4227>

REFERENCES

- Hodges B. Assessment in the post-psychometric era: learning to love the subjective and collective. *Med Teach*. 2013;35(7):564-568. doi:10.3109/0142159X.2013.789134
- Hodges B. Medical education and the maintenance of incompetence. *Med Teach*. 2006;28(8):690-696. doi:10.1080/01421590601102964
- Driessen E, van der Vleuten C, Schuwirth L, van Tartwijk J, Vermunt J. The use of qualitative research criteria for portfolio assessment as an alternative to reliability evaluation: a case study. *Med Educ*. 2005;39(2):214-220. doi:10.1111/j.1365-2929.2004.02059.x
- Schuwirth LW, van der Vleuten CP. A plea for new psychometric models in educational assessment. *Med Educ*. 2006;40(4):296-300. doi:10.1111/j.1365-2929.2006.02405.x
- Ten Cate O, Regehr G. The power of subjectivity in the assessment of medical trainees. *Acad Med*. 2019;94(3):333-337. doi:10.1097/ACM.0000000000002495
- Epstein RM, Hundert EM. Defining and assessing professional competence. *JAMA*. 2002;287(2):226-235. doi:10.1001/jama.287.2.226
- van der Vleuten CPM, Schuwirth LWT, Driessen EW, Govaerts MJB, Heeneman S. Twelve tips for programmatic assessment. *Med Teach*. 2015;37(7):641-646. doi:10.3109/0142159X.2014.973388
- van der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. *Med Educ*. 2005;39(3):309-317. doi:10.1111/j.1365-2929.2005.02094.x
- van der Vleuten CP, Schuwirth LW, Driessen EW, et al. A model for programmatic assessment fit for purpose. *Med Teach*. 2012;34(3):205-214. doi:10.3109/0142159X.2012.652239
- Schuwirth L, Van der Vleuten C. Programmatic assessment: from assessment of learning to assessment for learning. *Med Teach*. 2011;33(6):478-485. doi:10.3109/0142159X.2011.565828
- Schuwirth LW, van der Vleuten CP. Programmatic assessment and Kanes validity perspective. *Med Educ*. 2012;46(1):38-48. doi:10.1111/j.1365-2923.2011.04098.x
- Torre DM, Schuwirth L, van der Vleuten C. Theoretical considerations on programmatic assessment. *Med Teach*. 2020;42(2):213-220. doi:10.1080/0142159X.2019.1672863
- Heeneman S, de Jong LH, Dawson LJ, et al. Ottawa 2020 consensus statement for programmatic assessment-1. Agreement on the principles. *Med Teach*. 2021;43(10):1139-1148. doi:10.1080/0142159X.2021.1957088
- Driessen EW, van Tartwijk J, Govaerts M, Teunissen P, van der Vleuten CPM. The use of programmatic assessment in the clinical workplace: a Maastricht case report. *Med Teach*. 2012;34(3):226-231. doi:10.3109/0142159X.2012.652242
- Roberts C, Shadbolt N, Clark T, Simpson P. The reliability and validity of a portfolio designed as a programmatic assessment of performance in an integrated clinical placement. *BMC Med Educ*. 2014;14(1):197. doi:10.1186/1472-6920-14-197
- Dannefer EF. Beyond assessment of learning towards assessment for learning: educating tomorrows physicians. *Med Teach*. 2013;35(7):560-563. doi:10.3109/0142159X.2013.787141
- Ellaway RH, Kehoe A, Illing J. Critical realism and realist inquiry in medical education. *Acad Med*. 2020; Publish Ahead of Print;95(7):984-988. doi:10.1097/ACM.0000000000003232
- Roberts C, Khanna P, Lane AS, Reimann P, Schuwirth L. Exploring complexities in the reform of assessment practice: a critical realist perspective. *Adv Health Sci Educ*. 2021;26(5):1-17. doi:10.1007/s10459-021-10065-8
- Jorm C, Roberts C. Using complexity theory to guide medical school evaluations. *Acad Med*. 2018;93(3):399-405. doi:10.1097/ACM.0000000000001828
- Khanna P, Roberts C, Lane AS. Designing health professional education curricula using systems thinking perspectives. *BMC Med Educ*. 2021;21(1):1-8. doi:10.1186/s12909-020-02442-5
- Schuwirth L, Ash J. Assessing tomorrows learners: in competency-based education only a radically different holistic method of assessment will work. Six things we could forget. *Med Teach*. 2013;35(7):555-559. doi:10.3109/0142159X.2013.787140
- Hays R, Wellard R. In-training assessment in postgraduate training for general practice. *Med Educ*. 1998;32(5):507-513. doi:10.1046/j.1365-2923.1998.00231.x
- Wrigley W, van der Vleuten CP, Freeman A, Muijtjens A. A systemic framework for the progress test: strengths, constraints and issues: AMEE Guide No. 71. *Med Teach*. 2012;34(9):683-697. doi:10.3109/0142159X.2012.704437
- Wilkinson TJ, Tweed MJ, Egan TG, et al. Joining the dots: conditional pass and programmatic assessment enhances recognition of problems with professionalism and factors hampering student progress. *BMC Med Educ*. 2011;11(1):29. doi:10.1186/1472-6920-11-29
- Hodges B, Paul R, Ginsburg S, Members tOCG. Assessment of professionalism: from where have we come-to where are we going? An update from the Ottawa consensus group on the assessment of professionalism. *Med Teach*. 2019;41(3):249-255. doi:10.1080/0142159X.2018.1543862
- van Schaik S, Plant J, O'Sullivan P. Promoting selfdirected learning through portfolios in undergraduate medical education: the mentors' perspective. *Med Teach*. 2013;35(2):139-144. doi:10.3109/0142159X.2012.733832
- van Houten-Schat MA, Berkhout JJ, van Dijk N, Endedijk MD, Jaarsma ADC, Diemers AD. Self-regulated learning in the clinical context: a systematic review. *Med Educ*. 2018;52(10):1008-1015.
- Dijkstra J, Galbraith R, Hodges BD, et al. Expert validation of fit-for-purpose guidelines for designing programmes of assessment. *BMC Med Educ*. 2012;12(1):20. doi:10.1186/1472-6920-12-20
- Schuwirth L, Valentine N, Dilena P. An application of programmatic assessment for learning (PAL) system for general practice training. *GMS J Med Educ*. 2017;34(5):Doc56. doi:10.3205/zma001133
- Norcini J, Anderson MB, Bollela V, et al. 2018 Consensus framework for good assessment. *Med Teach*. 2018;40(11):1102-1109. doi:10.1080/0142159X.2018.1500016
- Burgess AW, McGregor DM, Mellis CM. Applying established guidelines to team-based learning programs in medical schools: a systematic review. *Acad Med*. 2014;89(4):678-688. doi:10.1097/ACM.000000000000162

32. Parmelee D, Michaelsen LK, Cook S, Hudes PD. Team-based learning: a practical guide: AMEE guide no. 65. *Med Teach*. 2012;34(5):e275-e287. doi:10.3109/0142159X.2012.651179
33. Burgess A, Bleasel J, Haq I, et al. Team-based learning (TBL) in the medical curriculum: better than PBL? *BMC Med Educ*. 2017;17(1):243. doi:10.1186/s12909-017-1068-z
34. Haidet P, Levine RE, Parmelee DX, et al. Perspective: guidelines for reporting team-based learning activities in the medical and health sciences education literature. *Acad Med*. 2012;87(3):292-299. doi:10.1097/ACM.0b013e318244759e
35. Danermark B, Ekström M, Karlsson JC. *Explaining society: Critical realism in the social sciences*. Routledge; 2019.
36. Bhaskar R. *A realist theory of science*. Routledge; 1975.
37. Archer M. *Structure, Agency and the Internal Conversation*. Cambridge University Press; 2003. doi:10.1017/CBO9781139087315.
38. Groff R. *Critical Realism, Post-positivism and the Possibility of Knowledge*. Routledge; 2004. doi:10.4324/9780203417270.
39. Bhaskar R, Danermark B, Price L. *Interdisciplinarity and Wellbeing: A Critical Realist General Theory of Interdisciplinarity*. Routledge; 2017. doi:10.4324/9781315177298.
40. Rutzou T. Realism. In: Kivisto P, ed. *The Cambridge Handbook of Social Theory: Volume 2: Volume II: Contemporary Theories and Issues*. Cambridge: Cambridge University Press; 2020.
41. Archer M. Realism and the problem of agency. *Alethia*. 2002;5(1):11-20. doi:10.1558/aleth.v5i1.11
42. Archer MS. *Being Human: The Problem of Agency*. Cambridge University Press; 2000. doi:10.1017/CBO9780511488733.
43. Priestley M, Edwards R, Priestley A, Miller K. Teacher agency in curriculum making: agents of change and spaces for manoeuvre. *Curric Inq*. 2015;42(2):191-214. doi:10.1111/j.1467-873X.2012.00588.x
44. Archer MS. Morphogenesis versus structuration: on combining structure and action. *Br J Sociol*. 1982;33(4):455-483. doi:10.2307/589357
45. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. *Qual Health Res*. 2016;26(13):1753-1760. doi:10.1177/1049732315617444
46. Meyer SB, Lunnay B. The application of abductive and retroductive inference for the design and analysis of theory-driven sociological research. *Sociol Res Online*. 2013;18(1):86-96. doi:10.5153/sro.2819
47. Fletcher AJ. Applying critical realism in qualitative research: methodology meets method. *Int J Soc Res Methodol*. 2017;20(2):181-194. doi:10.1080/13645579.2016.1144401
48. Cohen L, Manion L, Morrison K. *Research Methods in Education*. Routledge; 2013.
49. Sweller J. Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educ Psychol Rev*. 2010;22(2):123-138. doi:10.1007/s10648-010-9128-5
50. Senge PM. *The Leaders New Work: Building Learning Organisations*. Routledge; 2017.
51. Billett S. Conceptualising learning experiences: Contributions and mediations of the social, personal, and brute. *Mind Cult Act*. 2009;16(1):32-47. doi:10.1080/10749030802477317
52. Durkheim E. *The Division of Labour in Society (1893)*. Simon & Schuster; 1983.
53. Elder-Vass D. Re-examining Bhaskar's three ontological domains: the lessons from emergence. In: Lawson C, Latsis J, Martins N, eds. *Contributions to Social Ontology*. Routledge; 2016:160-176.
54. Archer MS. *Realist Social Theory: The Morphogenetic Approach*. Cambridge University Press; 1995. doi:10.1017/CBO9780511557675.
55. Archer MS. Routine, reflexivity, and realism. *Sociol Theory*. 2010;28:272-303. doi:10.1111/j.1467-9558.2010.01375.x
56. Schut S, Maggio LA, Heeneman S, van Tartwijk J, van der Vleuten C, Driessen E. Where the rubber meets the road—an integrative review of programmatic assessment in health care professions education. *Perspect Med Educ*. 2021;10(1):6-13. doi:10.1007/s40037-020-00625-w
57. Watling CJ, Ajjawi R, Bearman M. Approaching culture in medical education: three perspectives. *Med Educ*. 2020;54(4):289-295. doi:10.1111/medu.14037
58. Bearman M, Mahoney P, Tai J, Castanelli D, Watling C. Invoking culture in medical education research: a critical review and metaphor analysis. *Med Educ*. 2021;55(8):903-911. doi:10.1111/medu.14464
59. Varpio L, Aschenbrener C, Bates J. Tackling wicked problems: how theories of agency can provide new insights. *Med Educ*. 2017;51(4):353-365. doi:10.1111/medu.13160
60. Adie LE, Willis J, van der Kleij FM. Diverse perspectives on student agency in classroom assessment. *Aust Educ Res*. 2018;45(1):1-12. doi:10.1007/s13384-018-0262-2
61. Schut S, Driessen E, van Tartwijk J, van der Vleuten C, Heeneman S. Stakes in the eye of the beholder: an international study of learners perceptions within programmatic assessment. *Med Educ*. 2018;52(6):654-663. doi:10.1111/medu.13532
62. Schut S, van Tartwijk J, Driessen E, van der Vleuten C, Heeneman S. Understanding the influence of teacher-learner relationships on learners assessment perception. *Adv Health Sci Educ Theory Pract*. 2020;25(2):441-456. doi:10.1007/s10459-019-09935-z
63. Meeuwissen SNE, Stalmeijer RE, Govaerts M. Multiple-role mentoring: mentors conceptualisations, enactments and role conflicts. *Med Educ*. 2019;53(6):605-615. doi:10.1111/medu.13811
64. Schut S, van Tartwijk J, Driessen E, van der Vleuten C, Heeneman S. Understanding the influence of teacher-learner relationships on learners assessment perception. *Adv Health Sci Educ*. 2019;25(2):1-16. doi:10.1007/s10459-019-09935-z
65. Roberts C, Daly M, Held F, Lyle D. Social learning in a longitudinal integrated clinical placement. *Adv Health Sci Educ*. 2017;22(4):1011-1029. doi:10.1007/s10459-016-9740-3
66. Cilliers FJ, Schuwirth LW, Adendorff HJ, Herman N, van der Vleuten CP. The mechanism of impact of summative assessment on medical students learning. *Adv Health Sci Educ Theory Pract*. 2010;15(5):695-715. doi:10.1007/s10459-010-9232-9
67. Heeneman S, Oudkerk Pool A, Schuwirth LW, van der Vleuten CP, Driessen EW. The impact of programmatic assessment on student learning: theory versus practice. *Med Educ*. 2015;49(5):487-498. doi:10.1111/medu.12645
68. Watling CJ, Ginsburg S. Assessment, feedback and the alchemy of learning. *Med Educ*. 2019;53(1):76-85. doi:10.1111/medu.13645
69. Schuwirth LW, van der Vleuten CP. Current assessment in medical education: programmatic assessment. *J Appl Res Technol*. 2019;20:2-10.
70. Wright SR, Boyd VA, Ginsburg S. The hidden curriculum of compassionate care: can assessment drive compassion? *Acad Med*. 2019;94(8):1164-1169. doi:10.1097/ACM.0000000000002773
71. Sambell K, McDowell L. The construction of the hidden curriculum: messages and meanings in the assessment of student learning. *Assess Eval High Educ*. 1998;23(4):391-402. doi:10.1080/0260293980230406
72. Peters H, Zdravkovic M, João Costa M, et al. Twelve tips for enhancing student engagement. *Med Teach*. 2019;41(6):632-637. doi:10.1080/0142159X.2018.1459530
73. Milles LS, Hitzblech T, Drees S, Wurl W, Arends P, Peters H. Student engagement in medical education: a mixed-method study on medical students as module co-directors in curriculum development. *Med Teach*. 2019;41(10):1143-1150. doi:10.1080/0142159X.2019.1623385

How to cite this article: Roberts C, Khanna P, Bleasel J, et al. Student perspectives on programmatic assessment in a large medical programme: A critical realist analysis. *Med Educ*. 2022; 56(9):901-914. doi:10.1111/medu.14807