

**ASSESSMENT**

# Between trust and control: Teachers' assessment conceptualisations within programmatic assessment

Suzanne Schut<sup>1</sup>  | Sylvia Heeneman<sup>1,2</sup>  | Beth Bierer<sup>3</sup>  | Erik Driessen<sup>1</sup>  |  
Jan van Tartwijk<sup>4</sup>  | Cees van der Vleuten<sup>1</sup> 

<sup>1</sup>Department of Educational Development and Research, Faculty of Health, Medicine and Life Sciences, School of Health Professions Education, Maastricht University, Maastricht, the Netherlands

<sup>2</sup>Department of Pathology, Cardiovascular Research Institute Maastricht, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, the Netherlands

<sup>3</sup>Education Institute, Cleveland Clinic, Lerner College of Medicine, Case Western Reserve University, Cleveland, Ohio, USA

<sup>4</sup>Department of Education, Utrecht University, Utrecht, the Netherlands

**Correspondence**

Suzanne Schut, Department of Educational Development and Research, Maastricht University, Universiteitssingel 60, 6229 ER Maastricht, the Netherlands.  
Email: s.schut@maastrichtuniversity.nl

**Abstract**

**Objectives:** Programmatic assessment attempts to facilitate learning through individual assessments designed to be of low-stakes and used only for high-stake decisions when aggregated. In practice, low-stake assessments have yet to reach their potential as catalysts for learning. We explored how teachers conceptualise assessments within programmatic assessment and how they engage with learners in assessment relationships.

**Methods:** We used a constructivist grounded theory approach to explore teachers' assessment conceptualisations and assessment relationships in the context of programmatic assessment. We conducted 23 semi-structured interviews at two different graduate-entry medical training programmes following a purposeful sampling approach. Data collection and analysis were conducted iteratively until we reached theoretical sufficiency. We identified themes using a process of constant comparison.

**Results:** Results showed that teachers conceptualise low-stake assessments in three different ways: to stimulate and facilitate learning; to prepare learners for the next step, and to use as feedback to gauge the teacher's own effectiveness. Teachers intended to engage in and preserve safe, yet professional and productive working relationships with learners to enable assessment for learning when securing high-quality performance and achievement of standards. When teachers' assessment conceptualisations were more focused on accounting conceptions, this risked creating tension in the teacher-learner assessment relationship. Teachers struggled between taking control and allowing learners' independence.

**Conclusions:** Teachers believe programmatic assessment can have a positive impact on both teaching and student learning. However, teachers' conceptualisations of low-stake assessments are not focused solely on learning and also involve stakes for teachers. Sampling across different assessments and the introduction of progress committees were identified as important design features to support teachers and preserve the benefits of prolonged engagement in assessment relationships. These insights contribute to the design of effective implementations of programmatic assessment within the medical education context.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2020 The Authors. *Medical Education* published by Association for the Study of *Medical Education* and John Wiley & Sons Ltd

## 1 | INTRODUCTION

Interest in using assessment for learning is increasing in medical education and expectations of its benefits are high.<sup>1</sup> Programmatic assessment attempts to overcome the traditional dichotomy of assessment purposes as either formative or summative by proposing a continuum of assessment stakes.<sup>2,3</sup> This continuum of assessment stakes ranges from low (frequent assessments to benefit and support teachers and learners with information and feedback) to high (progress decisions based on the aggregation of assessment data). The primary goal of low-stake assessment is to support learners' progress. Thus, one low-stake assessment should have limited consequences for learners. When multiple low-stake assessments are aggregated, however, they can be used to inform high-stake performance decisions that have substantial consequences for learners.<sup>4</sup> In practice, learners often do not appreciate the value of low-stake assessments to guide their learning. Instead, they tend to focus on the potential summative consequences of low-stake assessments.<sup>5,6</sup> For this reason, using programmatic assessment to support learning remains challenging in practice.<sup>1,7,8</sup>

Teachers appear to play a particularly powerful role in fulfilling or undermining the learning potential of programmatic assessment.<sup>7</sup> Although many of the underlying principles of programmatic assessment may not be novel, the systematic approach to assessment and the continuum of assessment stakes with dual purposes fundamentally differ from traditional, summative approaches to assessment.<sup>9</sup> If teachers do not fully understand the meaning and purpose of assessment or do not agree with its underlying philosophy, low-stake assessments and their potential learning benefits are likely to become trivialised.<sup>4</sup> The complex and overlapping interplay of assessment purposes, such as in low-stake assessments, adds to the already complicated assessment processes.<sup>10,11</sup> Consequently, programmatic assessment may challenge teachers' conceptualisations of assessment.

Following the description of Thomson,<sup>12</sup> the concept of conceptions subsumes knowledge and beliefs into a singular construct and provides a framework for describing, in this context, teachers' overall perceptions and awareness of assessment. In the context of undergraduate teaching, Samuelowicz and Bain<sup>13</sup> confirmed conjectures in the literature that there is coherence between teachers' beliefs about teaching and learning processes (which range from those favouring the reproduction of knowledge and procedures to others favouring the construction and transformation of knowledge) and their assessment practices.<sup>13</sup> These authors warn that teachers may resist 'transformative' assessment methods for fundamental reasons and may not embrace innovation in assessment until they also shift their educational beliefs and values.<sup>14</sup> Furthermore, teachers' assessment conceptualisations are often informed by their personal assessment experiences rather than by educational theory or the institution's assessment policies.<sup>10,12</sup> These differences between beliefs and practices are especially likely to emerge when teachers encounter dual-purpose assessments,<sup>15</sup> such as the low-stake assessments used in programmatic assessment. For instance, teachers

may experience significant dilemmas when navigating between their supportive roles as they monitor and facilitate learners' development and their judgemental responsibilities as assessors of learners' performance and achievement.<sup>1,10,16,17</sup>

The perspective of teachers within programmatic assessment is a missing component in the medical education literature.<sup>18</sup> This qualitative study aims to address this gap by describing how teachers conceptualise assessment within programmatic assessment and exploring how teachers engage with learners in the context of programmatic assessment.

## 2 | METHODS

We used a constructivist grounded theory approach<sup>19,20</sup> to explore teachers' assessment conceptualisations and assessment relationships within programmatic assessment.

### 2.1 | Sample

An extreme case sampling strategy was employed to select unique research settings known to provide significant insights about programmatic assessment.<sup>21</sup> We selected research settings that required teachers to use low-stake assessment in contexts in which assessments have both low- and high-stake purposes. The inclusion criteria for these implementations were: (a) the use of low-stake assessment to provide information for learning; (b) the making of high-stake decisions on learners' progress based on the aggregation of those low-stake assessments, and (c) a long-term programmatic assessment implementation of at least 5 years. Based on previous research and suggestions by experts within the field, we selected two medical schools with graduate-entry medical programmes: the Physician-Clinical Investigator Programme at Maastricht University, the Netherlands (Setting A) and the Physician-Investigator Programme at the Cleveland Clinic Lerner College of Medicine at Case Western Reserve University, Cleveland, Ohio, USA (Setting B). These physician-investigator programmes aim to instil self-directed learning skills critical for the advancement of both biomedical research and clinical practice. Both programmes are competency-based, enrol small cohorts of students (<50 learners), and use programmatic assessment approaches to foster learning. The structure and characteristics of both programmes are shown in Table 1. Additionally, both programmes are described in detail elsewhere.<sup>5,22,23</sup>

We purposefully sampled participants using criterion and maximum variation sampling strategies. We invited teachers with formal responsibilities as assessors of low-stake assessment tasks for learners enrolled in the selected research sites or those whose main responsibilities involved providing feedback to guide students towards high-stake evaluation. Maximum variation was sought based on: (a) formal role in the programme (eg, tutor, coach, physician advisor/mentor, lecturer, coordinator, preceptor/supervisor); (b) type

of low-stake assessment (eg, standardised in-course tests, essays, [research] assignments, direct observations), and (c) variable lengths of relationships with learners (ranging from brief encounters to longitudinal relationships).

## 2.2 | Data collection and analysis

The lead investigator (SS) distributed an email to all selected participants describing the study and inviting them to participate voluntarily in semi-structured individual interviews on site. The research team designed an interview guide consisting of open-ended questions based on theoretical underpinnings of programmatic assessment and teachers' assessment conceptualisations. This interview guide included questions that asked participants to: (a) describe and reflect upon the concept of low-stake assessment within a programmatic assessment system; (b) discuss the roles and responsibilities of the teacher and learner in programmatic assessment; (c) reflect upon their interactions with learners in the context of programmatic assessment, and (d) articulate their values and beliefs about assessment and learning. Appendix S1 provides the initial interview guide. Although interviews focused upon assessment and assessment stakes within the implementation of programmatic assessment, participants were encouraged to reflect upon previous assessment experiences in order to help the research team fully understand teachers' assessment conceptualisations and experiences. All interviews were recorded and transcribed verbatim without direct identifiers.

Data collection and analyses were performed iteratively, allowing for necessary adaptations to interview questions and modifications of the sampling strategy for subsequent interviews.<sup>20,24</sup> The first four interviews were independently analysed by SS and SH using an open coding strategy with the aim of developing initial codes. Following each interview, SS and SH discussed the codes and relationships between codes. Based on these discussions, the initial codes were organised around key conceptual themes and sub-themes. Relationships amongst major categories were explored by examining and re-examining data. Initial codes evolved into conceptual codes, with examples and counter-examples. The research team (SS SH, BB, ED, JvT and CvdV) discussed the conceptual codes. To elaborate upon our preliminary analysis, we continued the use of theoretical sampling to gather additional perspectives about low-stake assessments in programmatic assessment. Specifically, we expanded our sample based on the teachers' experience in programmatic assessment and on teachers' backgrounds (teachers with basic science backgrounds versus clinicians). Data collection and analysis continued until theoretical sufficiency<sup>25</sup> was reached, meaning that we continued this data collection process until the analysis provided enough insight to understand teachers' assessment conceptualisations in the context of programmatic assessment. In total, 23 teachers participated in one-to-one, in-person interviews with the lead investigator (SS). Table 2 summarises the characteristics of these participants.

During data collection and analysis, SS created analytic memos and diagrams to ensure the process was logical and systematic. These memos and diagrams were discussed within the research team. Data were collected and analysed between December 2018

**TABLE 1** Characteristics of research settings

	Setting A	Setting B
Programme	Physician-Clinical Investigator Programme, Faculty of Health, Medicine and Life Science, Maastricht University, Maastricht, the Netherlands	Physician-Investigator Programme, Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, Cleveland, Ohio, USA
Duration	4-year graduate-entry	5-year graduate entry
Class size	50	32
Educational overview	PBL curriculum using a programmatic approach to assessment with the use of a portfolio and support of a mentor	PBL curriculum using a programmatic approach to assessment with the use of a portfolio and support of a physician advisor
Low-stake assessments	Knowledge (in- and end-of-block) tests, progress tests, clinical skills examinations, direct observations, field notes, clinical reasoning examinations, multi-source feedback rounds, critical appraisal of topics, essays, research seminars and presentations, peer and teacher feedback	Weekly SAQs and open-book CAPs, direct observations, OSCEs, journal club presentations, projects, research thesis, seminars and presentations, peer and teacher feedback
Feedback	A combination of narrative feedback on low-stake assessments and use of grades	Only narrative feedback on low-stake assessment, performance scores without pass or fail outcomes on SAQs, CAPs and OSCEs. No grades or class ranks
High-stake decision	Decisions made by a portfolio assessment committee based on learners' portfolios. Learners collect all assessment evidence and feedback into portfolios to monitor, analyse and reflect on strengths, weaknesses and progress	Decisions made by a promotion committee based on learners' portfolio essays. Learners compile portfolios to monitor, analyse and reflect on low-stake assessments and feedback with the aim of identifying strengths and targeting areas for improvement. The portfolio essay addresses the learner's progress and performance citing evidence from the portfolio

Abbreviations: CAP, concept appraisal; OSCE, objective structured clinical examination; PBL, problem-based learning; SAQ, short-answer question.

and May 2019. Ethical approval was obtained from the Dutch Association for Medical Education Ethical Review Board (NVMO-ERB ref. 2018.7.4) and the Cleveland Clinic's Institutional Review Board (IRB ref. 18-1516).

### 2.3 | Reflexivity

We acknowledge the roles that we, as researchers, played in collecting, analysing and interpreting these data. To help mitigate bias, we worked as a multidisciplinary research team. SS functioned as the lead researcher. SS has a background in educational sciences, works as a faculty member at one of the study sites, and had no direct involvement in the selected programme. ED and CvdV are experts in the field of medical education and assessment. Furthermore, CvdV is considered as one of the founding fathers of the theoretical model of programmatic assessment in medical education. SH has formal training and experience in the health sciences and BB has an equivalent background in teaching and research methods. Both SS and BB were involved as programme directors in the design and implementation of the selected programmes, as was CvdV as an expert. SH and BB had no direct contact with the participants during data collection. JvT is trained as a sociologist and is an expert in teacher education. JvT provided an outsider perspective to help thwart tunnel vision and confirmation bias, reviewed examples and counter-examples, and supported the process of code construction and data interpretation.

## 3 | RESULTS

The results showed that teachers conceptualise the purpose of low-stake assessment in three different, yet related ways: (a) to stimulate and facilitate learning; (b) to prepare learners for the next step, and (c) to use as feedback to gauge the teacher's own effectiveness. Consequently, these views influenced their engagement with learners when providing or discussing assessments. Results are presented through illustrative examples of verbatim quotes from participants and identified according to research site (A/B), interview sequence (1, 2, ...) and the participant's background (basic scientist or clinician).

### 3.1 | Conceptualisations of low-stake assessments

#### 3.1.1 | Stimulating and facilitating learning

Despite the differences in teachers' formal positions (eg, tutor, coach, physician advisor or mentor, course director, assessor, preceptor), we identified a shared primary conceptualisation of the purpose of low-stake assessments as being to stimulate and facilitate learning. This conception was influenced by the perceived minimal consequences of low-stake assessment. Statements like: 'learners

can't fail them,' 'they are not graded' and 'low-stake assessments are primarily about improving performance' were given by all participants when reflecting on the concept of low-stake assessments. The use of grades was strongly associated with high-stake assessments, and most participants did not regard assigning grades beneficial for student learning. Instead, grades were associated with the assessment purposes of ranking and comparing learners. To enable learners to use low-stake assessments for learning, teachers highlighted the importance of providing learners with narrative feedback in order to stimulate learning and facilitate improvement:

*The rank ordering of students is not that meaningful to me. [...] In this environment [programmatic assessment without the use of grades] there is not a fear of being incorrect as much, I think, and they [learners] are not trying to look smart in order to get rank order grades with this system.*

(B5, clinician)

At a programme level, the number of opportunities for collecting evidence on performance or improvement influenced teachers' assessment conceptualisations and opportunities for learning:

*There's only one chance in the programme, and so the progress committee will expect them [learners] to use it [the result of this assessment] in their portfolios, so that raises the stakes tremendously.*

(B7, basic scientist)

When the programme facilitated multiple low-stake assessments, teachers conceptualised their responsibility as being to support learners in discovering trends or patterns in assessment evidence, to stimulate reflection, and to enable learners' improvement plans for reaching learning goals and perceived potential. Furthermore, multiple low-stake assessments created better opportunities for teachers to provide learners with honest and constructive feedback because they perceived limited consequences:

*I think it's liberating in a lot of ways, because if you know that somebody can improve without being punished, there is no reason to not give them the information about something that is problematic. Whereas I think that in other settings, it feels like people get into the habit of highlighting things that learners are doing well and just being quiet about things that are problematic because 'I don't want anybody to get in trouble.'*

(B8, clinician)

#### 3.1.2 | Preparing learners for the next step

In addition to learning, teachers also thought of low-stake assessments as a way to prepare learners for high-stake assessments or for future practice. This assessment conceptualisation strongly

**TABLE 2** Participant characteristics (n = 23)

Characteristic	Setting	
	A	B
Programme	Physician-Clinical Investigator Programme, Faculty of Health, Medicine and Life Science, Maastricht University, the Netherlands	Physician-Investigator Programme, Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, Ohio, USA
Participants	9	14
Years of experience in programmatic assessment		
Beginner (<2 y)	4	3
Advanced (>5 y)	5	11
Gender		
Female	6	8
Male	3	6
Background		
PhD (basic sciences)	4	4
MD (physician)	5	10
Formal role <sup>a</sup> in assessment system		
Physician advisor or mentor Supports learners to interpret feedback, reflect on performance, monitor progress and develop portfolios	3	5
Tutor or coach Facilitator of PBL meetings, helps learners to construct subject matter-related objectives	4	5
Course or module director Responsible for design and delivery of courses or modules within curriculum	5	10
Preceptor or supervisor Responsible for guidance and supervision during clinical workplace-based assessment and learning	3	5
Portfolio, progress or promotion committee member Involved with portfolio reviews and shared responsibility for high-stake decision process	2	3

Abbreviations: MD, Doctor of Medicine; PBL, problem-based learning; PhD, Doctor of Philosophy.

<sup>a</sup>Most teachers performed multiple roles.

influenced how teachers facilitated learning: teachers thought a more directive approach was required to ensure learners were 'properly prepared.' What was considered important differed between basic scientists and clinicians.

Most teachers with teaching tasks related to the basic sciences within the curriculum emphasised assessment of knowledge. They regarded knowledge as fundamental for competence, and most believed learners should be able to pass a knowledge test:

*In my view these are important hurdles which they [learners] have to take at certain points. [...] If you are not capable of meeting those standards, you have insufficient knowledge and insights, which needs to have consequences.*

(A1, basic scientist)

Clinicians who participated in this study, however, tended to focus on overall clinical competence. Although knowledge testing

was considered important and often fundamental, gaps in knowledge were perceived as being easy for learners to remediate. According to many of the clinicians interviewed, these tests were considered as less important for preparing learners for 'real' clinical practice:

*I don't think they [knowledge tests] reflect what it means to be a physician.*

(B4, clinician)

Clinicians used low-stake assessment mainly to prepare learners for future practice:

*I think that is one of the ways that they improve their skills [by] preparing them and making sure they are optimised for [the] clinical years.*

(A3, clinician)

Exceptions were found when external, high-stake knowledge assessments were involved. All teachers understood that learners must pass high-stake assessments to meet either graduation or licensure requirements and considered preparing learners for such assessments an important responsibility, whether they considered the assessment meaningful or not:

*It's important that they see and practise with these types of questions and how they are styled, to prepare them for the way the National Board writes them, because those are really high stake.*

(B10, basic scientist)

### 3.1.3 | Low-stake assessments as feedback for teachers

Low-stake assessment also carried value for teaching practices and teachers themselves. Teachers conceptualised low-stake assessments as representing opportunities to diagnose learners' progress in acquiring learning objectives, to identify learners they thought required remediation, and to monitor learners' achievement of performance standards. Some teachers appreciated the reciprocal benefits low-stake assessment may have upon their personal and professional development, which stimulated a reflective attitude:

*It's a learning opportunity for the student, but, really, it's also a learning opportunity for me. It forces me to be reflective too, and think about what I'm doing, and what could be improved.*

(B4, clinician)

Teachers relied on low-stake assessments to inform them about their effectiveness. Teachers perceived learners' performances on low-stake assessments as explicit and direct indicators of their own performance, thereby making these assessments of higher stakes for teachers:

*For me it's [standardised knowledge test] a high-stake moment, and I'm relieved and very happy when students perform well on the test. It means I did a good job.*

(A1, basic scientist)

This observation also applied to clinical contexts, such as when teachers supervised individual learners during a clerkship or rotation:

*I know I do this. I am like: who did you have for your longitudinal clinic? And so this idea that this person has worked with me and this is where they are, I feel like it is a certain reflection of me and so then it feels like the stakes are higher as part of it, we are sending them out to the next preceptor and in the end, into the real world.*

(B8, clinician)

## 3.2 | Teachers' engagement with learners in assessment relationships

### 3.2.1 | Creating safe but productive relationships

When teachers' assessment conceptualisations focused on the use of assessment for learning, teachers indicated a strong need to create safe teacher-learner relationships, which they described using words such as 'care,' 'warmth,' 'accessible' and 'partnership.' Teachers were aware that learners often had different perceptions of assessment, and teachers took responsibility for orienting learners to the underlying philosophy of the assessment system. Teachers believed it was their responsibility to create a 'low-stake' learning environment in which learners could fail or make mistakes, and to use low-stake assessment to improve their performance. Teachers gained joy from partnering with learners and viewed the underlying philosophy of programmatic assessment as better aligned with real-life practice than traditional assessment approaches, thereby making their assessment practices with learners more meaningful and relevant:

*My job is not to be a gatekeeper anymore or keep students from graduating, but to help students be successful. My job now is: 'Are you getting better?' I feel much better about that role than [about] saying: 'You are done.'*

(B11, clinician)

Nevertheless, teachers focused on striking the right balance between maintaining safe learning environments and preserving productive working and assessment relationships with learners. This appeared to require a certain distance in the teacher-learner relationship. Teachers thought the relationship needed to be professional:

*They [learners] are not my friends or anything. I think it's important that I'm approachable, but there are certain boundaries; it needs to stay a professional relationship.*

(A19, clinician)

All teachers were explicit about not getting too close to or overly familiar with learners in the context of assessment; teachers wanted to minimise undue influences of their personal biases.

### 3.2.2 | Taking control versus allowing independence

Although teachers were explicit about their intention to allow learners to take responsibility for learning, almost all teachers believed that, in the end, they should control the assessment process. Teachers indicated that this was a natural consequence of their formal hierarchical position and their level of experience and expertise compared with those of learners. This need for control

was further augmented by teachers' high-stake responsibility concerning intended learning objectives and, in a clinical context, patient safety:

*But I am in control. I mean, I am, you know it is my responsibility to make sure they are learning. [...] There are things that need to be done and that they have to learn. If I left it to them... who knows? So, I really need to be able to control it. [...] You have to make sure that someone is skilled in doing something before you allow them to do it.*  
(B4, clinician)

Novice teachers in programmatic assessment desired more control of assessment processes than experienced teachers. Those with limited experience with programmatic assessment voiced uncertainties about their knowledge and proficiency with programme demands and the effectiveness of the assessment system as a whole. As a result, they perceived a high level of pressure on the quality of their guidance and support and feared that learners might be penalised as a result of their lack of experience with programmatic assessment. More experienced teachers, who explicitly valued learners' autonomy, seemed more comfortable with allowing learners to take additional control over assessment processes. This was strongly influenced by teachers' beliefs in learners' abilities and competencies:

*I think it's important to adapt to individual student needs [...], the need for independence grows over time.*  
(A21, basic scientist)

### 3.2.3 | Conflicts in assessment relationships

The potential conflicts teachers were able to perceive in teacher-learner assessment relationships seemed most likely to occur when teachers interacted with problematic or underperforming learners. Teachers voiced discomfort about providing learners with constructive or critical feedback and worried about preserving relationships:

*I think that discomfort with 'I'm the one that is going to have to identify that they haven't done what they're supposed to do,' is not why I chose to be a medical educator.*  
(B8, clinician)

Furthermore, teachers attributed their discomfort to the perceived need to provide more supervision for struggling learners, such as additional meetings and more extensive feedback. This raised concerns about what would actually be assessed in the final high-stake decision on learner performance: the teacher's mentoring and feedback skills or the learner's performance and progress?

A productive working relationship with struggling learners was easier to maintain when progress committees assumed responsibility for high-stake performance decisions and functioned as external parties to teacher-learner assessment relationships. Moreover,

teachers conceptualised assessment decisions within a programmatic approach as a shared responsibility, which most perceived as representing a positive change from their previous assessment experiences:

*You need more people. We kind of correct each other's perspectives on things and offer things that are helpful. That also makes it safer for the student. [...] The wisdom of several is better than the wisdom of some.*  
(B11, clinician)

## 4 | DISCUSSION

The aims of this study were to describe teachers' assessment conceptualisations within programmatic assessment and to explore how teachers engage with learners in the context of programmatic assessment. The findings showed that teachers conceptualise low-stake assessments in three ways, which are not solely focused on student learning. These conceptualisations give rise to potential tensions in the teacher-learner assessment relationship, which we will now discuss in the light of the existing literature.

The assessment continuum within programmatic assessment theoretically flows from one extreme (the 'learning conception of assessment') to the opposite extreme (the 'accounting conception of assessment') yet holds a dual purpose in each single low-stake assessment.<sup>2,3</sup> Most teachers focused on a learning conception of low-stake assessment. However, when 'learning' was conceived as preparing learners for high-stake assessment and when teachers emphasised teachers' accountability, teachers' assessment conceptualisations actually moved towards the accounting end of the continuum and carried a more directing and controlling tone. Such conceptualisations risk teaching to the test, whether it is considered meaningful or not, especially when external high-stake assessments are involved. This adverse impact of external assessment has been described by Stiggins,<sup>26</sup> who notes that centralised assessment for accountability purposes cannot meet the instructional information needs of individual teachers and may run the risk of trivialising their assessment practices. Although the results showed that the implementation of programmatic assessment could enable a shift in teachers' focus on the acquisition of the knowledge and skills necessary for learners to pass a test to a focus on continuous professional development and clinical competence, high-stake and especially standardised examinations could impede the occurrence of this shift.

The results of this study further showed that the stakes of low-stake assessment are just as much involved for teachers when teachers gauge their effectiveness based on learners' performance and progression. This may explain why so many teachers desire to control assessment processes to ensure high-quality learner performance and achievement of performance standards. Teachers in our study were aware of the learner's position of dependency and expressed a paradox when describing teacher-learner assessment relationships.

The valuing of teacher-learner partnerships, learner independence and learner self-regulation abilities did not appear to be sufficient for teachers to lessen their control of assessment processes. Teachers admitted that they empowered learners to take more control over assessment processes only when the learner's performance or competence aligned with the teacher's perceptions of 'good' practice or established criteria. This unilateral determination by teachers of what constitutes good practice seems at odds with the objective of self-regulation<sup>27-29</sup> and could work counterproductively when assessment is intended to be used for learning. Furthermore, this need for control on the part of the teacher may explain why learners so often fail to perceive low-stake assessments as being truly of low stakes and beneficial for their learning.<sup>5-7,30</sup> The importance of learner agency, defined as the learner's ability to act, control and make choices within the learning and assessment environment, is voiced by many scholars.<sup>1,31,32</sup> Moreover, learners themselves have voiced the importance of agency to enable the potential of using assessment for their learning.<sup>7</sup> Here too lingers the tension between trust and control. If we want learners to enjoy a safe low-stake environment in order to facilitate assessment for learning, then we should focus on creating supportive low-stake environments for teachers as well. Stakes are involved for both teachers and learners, and they are clearly not as straightforward as the low consequence of a single assessment.

The results also identified two important programmatic assessment design features that seemed to support teachers' use of low-stake assessment for learning: (a) the use of multiple low-stake assessments, especially those without the use of grades, and (b) the implementation of progress committees, which introduces an independent third party into the assessment relationship. First, the principle of using multiple low-stake assessments and assessors enabled teachers to provide more honest and critical feedback to learners, which, in light of medical education's 'failure to fail'<sup>33</sup> is a promising design feature of the programmatic assessment approach. Previous research has shown that both progress committees and learners rate the quality of low-stake assessment evidence more highly when assessment evidence originates from different contexts and sources.<sup>34</sup> Thus, the number of opportunities for collecting assessment evidence provided by the programme strongly influences the perceptions of assessment stakes and learning value for the multiple stakeholders involved.<sup>7,34</sup> Furthermore, the emphasis on narrative feedback, as opposed to the use of grades, was perceived as a key design factor to enable assessment for learning because such feedback emphasises mastery and progress instead of comparison, ranking and competition. The risks associated with the use of grades and the importance of narrative feedback to promote learning have been highlighted by many others.<sup>1,30,35-37</sup> Second, teachers enjoyed partnering with learners in the context of assessment and invested in engaging in productive working relationships with learners. Although for some teachers the dual purpose of low-stake assessment may continue to represent an unhappy marriage, our results showed that a role conflict is not necessary. Similar findings emerged in a study on multiple-role mentoring in programmatic assessment.<sup>38</sup> Conflicts in our study were reported only in relation to struggling and underperforming learners.

The implementation of independent progress committees, also in use as clinical competency committees,<sup>39</sup> created opportunities for teachers to deal with this conflict more easily when preserving a productive teacher-learner relationship in an assessment context.

Our findings may benefit other implementations of programmatic assessment. Teachers worry about disadvantaging learners with assessment. A progress committee, when organised well, provides support, expertise and, more importantly, a safety net for teachers involved in programmatic assessment. Failure of a student becomes a collective responsibility and learners' careers do not rest on decisions made by individuals or on limited snapshots. This seems to take some of the pressure from teachers and allows them to provide more honest constructive feedback or to raise concerns when preserving the benefits of prolonged engagement.<sup>4</sup> Furthermore, participating in progress committees seems to contribute to teachers' shared understanding concerning assessment objectives and benefits teachers' professional development in their roles as assessors in programmatic assessment.

The different conceptualisations of low-stake assessment indicate that teachers are likely to hold varying beliefs about assessment, at least some of which may be contrary to the underlying assessment philosophy advocated by its developers. As students encounter many different teachers during medical training, it is likely that they will encounter teachers with different values or beliefs about assessment that do not align with the intentions and assessment methods used in a programme. This risks the possibility that learners will have experiences of irreconcilable assessment objectives or messages and lead them to follow a cynical 'give them what they want' approach,<sup>13</sup> which would hinder a meaningful uptake of assessment for learning. Moreover, teachers may resist or dismiss innovative assessment methods and complex dual-purpose systems, like programmatic assessment, if these methods and approaches do not align with their fundamental beliefs about education and teaching.<sup>13</sup> Faculty development should focus on the underlying principles of programmatic assessment and teachers' assessment conceptualisations as these may affect their assessment practices when engaging with learners in assessment relationships. Future research is needed to better understand the interaction between conceptualisations and assessment practices and when and how teachers use different approaches in practice. Observational research could provide additional insights into the interactions between teachers and students, what teachers actually do in practice, and how this affects learners' perceptions of programmatic assessment.

#### 4.1 | Limitations

Our findings should be considered in the light of a number of limitations. First, this study included two unique implementations of programmatic assessment (ie, a small cohort size, using criteria that selected both highly motivated learners and teachers). We purposefully investigated these so-called extreme cases in view of their ability to provide insight into the mechanisms underlying implementations,



which can serve as lessons to guide future research and practice.<sup>19</sup> Second, assessment is a complex interaction of learner, task, teacher and context characteristics,<sup>40</sup> which makes generalisations to other contexts challenging.<sup>41</sup> Teachers' roles and responsibilities can vary amongst programmes, institutions and cultural contexts. By purposefully seeking maximum variation in formal roles and assessment responsibilities, we focused on the underlying conceptualisation of teaching and assessment in programmatic assessment. Third, this study explored teachers' perceptions of their reality. There may be differences between what teachers report they believe and intend to do versus what they actually believe and do. Finally, we may have introduced selection bias as we recruited teachers who volunteered to participate in response to a direct solicitation email.

## 5 | CONCLUSIONS

Given the influence and importance of assessment in medical education, we need to design assessment programmes that have positive impacts on both teaching and learning. This study shows that teachers believe that programmatic assessment can engender such an impact. However, teachers' conceptualisations of low-stake assessments are not focused solely on learning. The use of assessment to monitor teaching effectiveness may create tension in teachers' assessment practices and the teacher-learner assessment relationship. Understanding the position of teachers' assessments conceptualisations represents a step towards influencing and perhaps changing those conceptualisations to align with assessment for learning practices. Sampling across different assessments and assessors and the introduction of progress committees were identified as important design features of programmatic assessment that support teachers in using assessment to benefit learning, when preserving the benefits of prolonged engagement. These insights may serve to guide further practical developments and contribute to the design of more effective and efficient programmes of assessment and their implementation within the medical education context.

### AUTHOR CONTRIBUTIONS

SS conducted the data collection, led the data analysis and drafted the manuscript. SH participated in independent data analysis. ED, JvT and CvdV checked the data analysis. SH, BB, ED, JvT and CvdV contributed to the critical revision of the paper. All authors (SS, SH, BB, ED, JvT and CvdV) contributed to the conceptualisation and design of the study as well as to the data interpretation, approved the final version of the manuscript, and have agreed to be accountable for all aspects of this work.

### ACKNOWLEDGEMENTS

The authors would like to acknowledge and appreciate the legacy of Dr Elaine Dannefer, whose contributions continue to impact the assessment system at the Cleveland Clinic Lerner College of Medicine.

### CONFLICTS OF INTEREST

None.

### ETHICAL APPROVAL

Ethical approval was obtained from the Dutch Association for Medical Education Ethical Review Board (NVMO-ERB ref. 2018.7.4) and the Cleveland Clinic's Institutional Review Board (IRB ref. 18-1516).

### ORCID

Suzanne Schut  <https://orcid.org/0000-0002-8298-399X>

Sylvia Heeneman  <https://orcid.org/0000-0002-6103-8075>

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

Erik Driessen  <https://orcid.org/0000-0001-8115-261X>

Jan van Tartwijk  <https://orcid.org/0000-0001-6804-4163>

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

### REFERENCES

1. Watling CJ, Ginsburg S. Assessment, feedback and the alchemy of learning. *Med Educ*. 2019;53(1):76-85.
2. Van der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. *Med Educ*. 2005;39(3):309-317.
3. Schuwirth LW, van der Vleuten CP. Programmatic assessment: from assessment of learning to assessment for learning. *Med Teach*. 2011;33(6):478-485.
4. 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.
5. 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.
6. Bok HG, Teunissen PW, Favier RP, et al. Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Med Educ*. 2013;13:123.
7. 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.
8. Harrison C, Wass V. The challenge of changing to an assessment for learning culture. *Med Educ*. 2016;50(7):704-706.
9. Uijtdehaage S, Schuwirth LWT. Assuring the quality of programmatic assessment: moving beyond psychometrics. *Perspect Med Educ*. 2018;7(6):350-351.
10. Looney A, Cumming J, van der Kleij F, Harris K. Reconceptualising the role of teachers as assessors: teacher assessment identity. *Assess Educ*. 2017;25(5):442-467.
11. Black P, William D. Assessment and classroom learning. *Assess Educ*. 2006;5(1):7-74.
12. Thomson AG. Teachers' beliefs and conceptions: a synthesis of the research. In: Grouws D, ed. *National Council of Teachers of Mathematics Handbook of Research on Mathematics Teaching and Learning*. Reston, VA: Information Age Publishing; 1992:127-146.
13. Samuelowicz K, Bain JD. Identifying academics' orientations to assessment practice. *High Educ*. 2002;43(2):173-201.
14. Samuelowicz K, Bain JD. Revisiting academics' beliefs about teaching and learning. *High Educ*. 2001;41(3):299-325.
15. Brown GTL, Lake R, Matters G. Queensland teachers' conceptions of assessment: the impact of policy priorities on teacher attitudes. *Teach Educ*. 2011;27(1):210-220.
16. Rea-Dickins P. Understanding teachers as agents of assessment. *Lang Test*. 2004;21(3):249-258.

17. Wiliam D. *Embedded Formative Assessment*. Bloomington, IN: Solution Tree Press; 2011.
18. Acai A, Li SA, Sherbino J, Chan TM. Attending emergency physicians' perceptions of a programmatic workplace-based assessment system: the McMaster Modular Assessment Program (McMAP). *Teach Learn Med*. 2019;31(4):434-444.
19. Corbin J, Strauss A. *Basics of Qualitative Research, 3rd edn. Techniques and Procedures for Developing Grounded Theory*. Thousand Oaks, CA: SAGE Publications Ltd; 2008.
20. Charmaz K. *Constructing Grounded Theory. A Practical Guide through Qualitative Analysis*. Thousand Oaks, CA. SAGE Publications Ltd; 2006.
21. Creswell JW. *Educational Research: Planning, Conducting and Evaluating Quantitative and Qualitative Research*. Harlow, UK: Pearson Education; 2014.
22. Dannefer EF, Henson LC. The portfolio approach to competency-based assessment at the Cleveland Clinic Lerner College of Medicine. *Acad Med*. 2007;82(5):493-502.
23. Dannefer EF. Beyond assessment of learning toward assessment for learning: educating tomorrow's physicians. *Med Teach*. 2013;35(7):560-563.
24. Watling CJ, Lingard L. Grounded theory in medical education research: AMEE Guide No. 70. *Med Teach*. 2012;34(10):850-861.
25. Dey I. *Grounding Grounded Theory: Guidelines for Qualitative Inquiry*. San Diego, CA: Academic Press; 1999.
26. Stiggins R. Two disciplines of educational assessment. *Meas Eval Counsel Dev*. 1993;26(1):93-104.
27. Pratto F. On power and empowerment. *Br J Soc Psychol*. 2016;55(1):1-20.
28. Heron J. Assessment revisited. In: Boud D, ed. *Developing Student Autonomy in Learning*, 2nd edn. London, UK: Kogan Page; 1981:55-68.
29. Schut S, Driessen E. Setting decision-making criteria: is medical education ready for shared decision making? *Med Educ*. 2019;53(4):324-326.
30. Harrison CJ, Konings KD, Dannefer EF, Schuwirth LW, Wass V, van der Vleuten CP. Factors influencing students' receptivity to formative feedback emerging from different assessment cultures. *Perspect Med Educ*. 2016;5(5):276-284.
31. Harrison CJ, Konings KD, Schuwirth L, Wass V, van der Vleuten C. Barriers to the uptake and use of feedback in the context of summative assessment. *Adv Health Sci Educ Theory Pract*. 2015;20(1):229-245.
32. Cilliers FJ, Schuwirth LW, van der Vleuten CP. A model of the pre-assessment learning effects of assessment is operational in an undergraduate clinical context. *BMC Med Educ*. 2012;12:9.
33. Yepes-Rios M, Dudek N, Duboyce R, Curtis J, Allard RJ, Varpio L. The failure to fail underperforming trainees in health professions education: a BEME systematic review: BEME Guide No. 42. *Med Teach*. 2016;38(11):1092-1099.
34. Dannefer EF, Bierer SB, Gladding SP. Evidence within a portfolio-based assessment program: what do medical students select to document their performance? *Med Teach*. 2012;34(3):215-220.
35. Konopasek L, Norcini J, Krupat E. Focusing on the formative: building an assessment system aimed at student growth and development. *Acad Med*. 2016;91(11):1492-1497.
36. Lefroy J, Hawarden A, Gay SP, McKinley RK, Cleland J. Grades in formative workplace-based assessment: a study of what works for whom and why. *Med Educ*. 2015;49(3):307-320.
37. Telio S, Regehr G, Ajjawi R. Feedback and the educational alliance: examining credibility judgements and their consequences. *Med Educ*. 2016;50(9):933-942.
38. Meeuwissen SNE, Stalmeijer RE, Govaerts M. Multiple-role mentoring: mentors' conceptualisations, enactments and role conflicts. *Med Educ*. 2019;53(6):605-615.
39. Colbert CY, Dannefer EF, French JC. Clinical competency committees and assessment: changing the conversation in graduate medical education. *J Grad Med Educ*. 2015;7(2):162-165.
40. Gipps CV. *Beyond Testing: Towards a Theory of Educational Assessment*. London, UK: Falmer Press; 1994.
41. Black P, Wiliam D. Lessons from around the world: how policies, politics and cultures constrain and afford assessment practices. *Curric J*. 2005;16(2):249-261.

#### SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Schut S, Heeneman S, Bierer B, Driessen E, van Tartwijk J, van der Vleuten C. Between trust and control: Teachers' assessment conceptualisations and relationships within programmatic assessment. *Med Educ*. 2020;54:528–537. <https://doi.org/10.1111/medu.14075>