

The burden of proof

Citation for published version (APA):

Schut, S. (2021). *The burden of proof: Agency and Accountability in Programmatic Assessment*. [Doctoral Thesis, Maastricht University]. Maastricht University. <https://doi.org/10.26481/dis.20211209ss>

Document status and date:

Published: 01/01/2021

DOI:

[10.26481/dis.20211209ss](https://doi.org/10.26481/dis.20211209ss)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

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

[Link to publication](#)

General rights

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

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

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

www.umlib.nl/taverne-license

Take down policy

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

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.



THE BURDEN OF PROOF

AGENCY AND ACCOUNTABILITY IN PROGRAMMATIC ASSESSMENT

SUZANNE SCHUT



THE BURDEN OF PROOF

AGENCY AND ACCOUNTABILITY IN PROGRAMMATIC ASSESSMENT

SUZANNE SCHUT

The research reported here was carried out at



in the School of Health Professions Education



© Suzanne Schut, 2021, Maastricht, the Netherlands.

ISBN: 978-94-6416-850-1

Design inspired by: Kunstmuseum, The Hague, H.P. Berlage and
Sol LeWitt Wall Drawing #373

Layout: PubliSS | www.publiss.nl

Printed by: Ridderprint | www.ridderprint.nl

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, by photocopying, recording, or otherwise, without the prior written permission of the author.

THE BURDEN OF PROOF

Agency and Accountability in Programmatic Assessment

DISSERTATION

to obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus,
Prof. dr. Rianne M. Letschert
in accordance with the decision of the Board of Deans,
to be defended in public
on Thursday 9th December 2021, at 14.00 hours

by

Suzanne Schut

Supervisors

Prof. dr. S. Heeneman
Prof. dr. E. Driessen
Prof. dr. J. van Tartwijk, Utrecht University
Prof. dr. C. van der Vleuten

Assessment committee

Prof. dr. P. Teunissen (chair)
Dr. M. Govaerts
Dr. C. Harrison, University of Central Lancashire
Prof. dr. R. Rennenberg
Prof. dr. M. van der Schaaf, Utrecht University

'Heart and humor and humility will lighten up your heavy load.'

Joni Mitchell

Table of Contents

Chapter 1	Introduction	9
Chapter 2	Stakes in the eye of the beholder: An International Study of Learners' Perceptions within Programmatic Assessment <i>Medical Education (2018)</i> , 52(6), 654-663	27
Chapter 3	Between trust and control: Teachers' assessment conceptualisations within programmatic assessment <i>Medical Education (2020)</i> , 54(6), 528-537	43
Chapter 4	Understanding the influence of teacher-learner relationships on learners' assessment perception. <i>Advances in Health Science Education: Theory and Practice (2020)</i> , 25(2), 441-456	63
Chapter 5	Setting decision-making criteria: is medical education ready for shared decision making? <i>Medical Education (2019)</i> , 53(4), 324-326	85
Chapter 6	Meaningful mapping of remediation in longitudinal and developmental assessment models <i>Medical Education (2020)</i> , 54(10), 866-868	89
Chapter 7	Where the rubber meets the road: An integrative review of programmatic assessment in health care professions education <i>Perspectives in Medical Education (2021)</i> , 10(1), 6-13	95
Chapter 8	Discussion	117
Addendum	Summary	142
	Impact paragraph	145
	Acknowledgments	148
	About the author	150
	SHE Dissertation series	151



CHAPTER 1

Introduction

INTRODUCTION

Assessment is often a burden for those involved in education. It is strongly associated with undesirable and unintended consequences for teaching and learning: assessment can lead to surface learning approaches, strategic choices, compliance behaviour, teaching and learning to the test, anxiety, stress, and fear of failing.^{e.g.1-5} Furthermore, assessment can transmit a message to students of what they can and cannot succeed in doing: some students have a majority of success experiences, passing most of their assessments and by doing so they build confidence for future assessments. Others have more experience in failing assessments and when such experiences prevail, they may even come to perceive assessment as a way of showing how inadequate they are as learners.²

Although assessment is argued to have a powerful influence on learning in education, the relationship between assessment and learning is complex and often poorly understood.^{6, 7} One of the explanations for this is that, when it comes to assessment, the explicit acknowledgment of the range of beliefs of what is real and knowable and how people come to know and understand has been largely overlooked.^{8, 9} Consequently, a desired influence of assessment on learning can be undermined due to naïve or unintended contradictions between beliefs about learning and how learning is justified in assessment practices. Conflicting assumptions of learning and assessment can lead to confusing and poorly supported practices with multiple possible interpretations of the impact or results of assessment.⁸ In order to prevent assessment to influence learning in an undesirable way, one must engage with the more normative question of what assessment and learning are supposedly for. As long as we are not more assured that what occurs under the banner of assessment and whether or not it influences learning and teaching in desirable ways, assessment is likely to remain a burden for those teaching and learning and for those being assessed.²

Whether or not we can find a more desirable interaction between assessment and learning and overcome the undesirable influences of assessment on learning forms the overall question of this dissertation. I¹ aim to address the complex interplay between assessment and learning with a series of studies¹¹ exploring the learning potential of a whole-system approach assessment in the context of competency-based medical education (CBME). It is the purpose of this introductory chapter to explain how discourse and conceptualisation of assessment in medical education evolved, how this inspired this dissertation, and to make my own assumptions, stances and frameworks explicit as these have shaped the content of this dissertation.

I This dissertation is accomplished under supervision and in collaboration with a research team, however, to acknowledge my final responsibility for this project I will use the first-person pronoun throughout the Introduction and General Discussion chapters.

II This dissertation is a collection of published journal articles and since every chapter is written to be read and understood on its own, some repetition across chapters is inevitable

Assessment in competency-based medical education

According to the Lancet Commission of Education for Health Professionals, there are about 2420 medical schools and 467 schools or departments around the world, educating about 1 million new doctors and other health care professional every year.¹⁰ A defining feature of competency-based medical education (CBME) is its focus on outcomes formulated as specific competencies dominate in postgraduate training. Since the introduction of the CanMEDS framework by the Royal College of Physicians and Surgeons of Canada¹¹ and the Outcome Project of the Accreditation Council for Graduate Medical Education in the USA,¹² CBME has become widely used terminology met with both enthusiasm and criticism, partly due to the varied interpretations.¹³ Ten Cate¹³ argues that CBME can be generally defined as ‘education for the medical profession that is targeted at a fixed level of proficiency in one or more medical competencies’.^(p.3) Although what is considered competent depends on the social and cultural context apparent in the dominant discourse at hand,¹⁴ it is generally acknowledged that licenced practitioners are expected to deliver high-quality health care and to be committed to continuous professional development and ongoing advancement of their field.¹⁵ Following this acknowledgement, the purpose of CBME is to *ensure* quality of competence while simultaneously *equip* its graduates for ongoing learning after formal education.

A general way in which one could define assessment is as any method to obtain information to determine the merit, value, state or worth of something, and to draw inferences about any type of object or subject based on the information obtained.⁴ Although interest in assessment and the role it plays in societies goes back for centuries, there have been some recent and significant developments in the literature that address assessment and CBME. Reliance in the assessment of medical schools on what became to be considered as too subjective methods, like essays and oral examinations, led to the development of a large array of psychometrically robust assessment tools.¹⁶ Psychometric approaches to assessment conceptualise competence as a combination of personality traits, like knowledge, skills, and attitudes and have the assumption that these could and should be measured independently.¹⁷ They originate from positivist stances and are associated with a discourse using words as: reliability, data, generalisability, checklist, and cut-of-points.¹⁴ In this paradigm, assessment is conceptualised as a measurement problem.¹⁷ However, trying to break down clinical performance into smaller, measurable units, as these approaches to assessment do, was said to trivialise the task and risk the validity of competency-based assessment.¹⁸ Amongst others, Govaerts and Van der Vleuten⁹ advocated that we need to extend our focus from such a predictive and deterministic framework towards constructivist-interpretivist approaches in order to capture and understand the contextualised learning in a longitudinal manner that is claimed to be more aligned with the philosophical position and purpose of competency-based medical education and performance in work settings.⁹ Assessment of CBME is argued to require a holistic longitudinal assessment approach, depending more heavily on professional, expert and thus interpretivist judgements of learning outcomes.^{9, 17, 19} In this paradigm, assessment is conceptualised as human judgement problem.¹⁷

Such developments in the literature led to a reconceptualisation of competencies and its assessment in CBME. First, instead of conceptualising assessment as a measurement problem which aims to produce structured, standardised and objective assessments and values concepts such as reliability and construct validity, assessment is proposed as an integrated whole system problem which embraces and values the role of human judgement with the use of expertise, good support and a need for literacy.¹⁷ Second, labelling someone ‘competent’ at a certain point in time has led to unintended consequences of the notion of competence as something one can check off and requires no further development needed.²⁰⁻²² This is due to increased emphasis on the important role played by contextual factors in the ability to perform any task, the considerable variability of performance on such tasks situated in the context, and the decay of knowledge and skills when no longer actively used because one was deemed ‘good enough’. In other words, labelling someone competent mistakes the reality of the need for continuous development and improvement.^{20, 21} Moreover, such notion of competence runs counter to perspectives on adaptive expertise,²³ which entails a continuous reinvestigation of effort and attention into better understanding and advancing the domain of practices for the sake of ongoing performance improvement.^{23, 24}

If the purpose of competency-based assessment is indeed thought of as to not only qualify licensed practitioners and ensure these qualifications are met given certain social and cultural contextual standards, but also to sufficiently equip graduates to learn in situations in which teachers and assessors are no longer present and to focus or even self-direct their effort and attention as they are expected to proceed with learning after formal graduation, the assessment programme needs to fulfil a double duty or purpose. The overall question which was posed in the beginning of this chapter concerned whether or not assessment can have a more desirable interaction with learning and whether or not we can overcome the undesirable or negative influences of assessment on learning. Could such a double purpose of the assessment programme solve this central problem? Or might assessment programmes be tainted by the accountability discourse that is often so strongly associated with and might even be considered to dominate assessment? To unravel this problem further, I will first shortly review different purposes of assessment from a more general perspective, the formative and summative purposes of assessment.

The purpose of assessment

For decades, researchers and educators have been debating the importance or unfeasibility of the distinction between the formative and summative purpose of assessment.^{e.g.,20, 25-29} With summative assessment the focus is on making judgements on the outcome of learning processes. These assessments occur at defined points, often at the end of a course or training phase, and are used for selection, certification and accountability purposes. As a result, they tend to shape teaching, learning and the curriculum, and due to the high-stakes consequences of these assessments, mechanisms are put in place to ensure accuracy, consistency, and reliability. The formative purpose of assessment originates from the context of programme evaluation. Formative evaluation was coined by Scriven³⁰ and used to describe the role that evaluation could

play in the on-going improvement of the curriculum, for instance by brief tests used by teachers and learners as aids in the learning process. The purpose of formative assessment is to provide feedback and correctives at each stage in the teacher-learning process. Formative assessment is said to work best when it is embedded in the teaching process and/or authentic workflow, provides specific and actionable feedback, is ongoing, and is timely.³¹ Crooks³² stressed that, although formative assessment has the potential power to influence learning, too often the use of assessment for summative purposes gets in the way. Similarly, Bloom³³ already worried about conflicting purposes and argued that ‘while formative tests may be graded and used as part of the judging and classificatory function of evaluation (a summative function), we see much more effective use of formative evaluation if it is separated from the grading process and used primarily as an aid to teaching’.^(p.48) However, the contrary case was also made. Taras²⁹ argues not only that we should not, but that we cannot separate formative and summative assessment. For assessment to be formative, it requires feedback, which indicates the existence of a ‘gap’ between the actual level of the work being assessed and the desired level. According to this view it is not even possible for assessment to be uniquely formative without a preceding summative judgement. Assessment can only be uniquely summative when the assessment process stops at the judgement moment.²⁹ Eva et al.,²⁰ commemorate that, although whether or not it is believed assessments can fulfil a dual purpose, the distinction is important in that it helps to make sense of the compromises that are appropriate in each, however, using this distinction as a rule would be detrimental. They highlight different risks that are involved, of which a fundamental one is that ‘the assumption that the duality of purpose cannot be achieved, simply mistakes the reality of the learner’s experience’.^{20 (p.903)} Even if assessment is intended to be purely formative, learners’ perception is likely to have aspects of summative judgement when performing a task in which some part of their identity is involved.³⁴ Furthermore, it will have an impact on whether or not the assessor thinks the learner is worth the effort of providing feedback and the same goes for learners and whether or not they think further improvement is worth their while.³⁴

Assessment for learning has begun to take a place on the agenda of institutions and there has been a flourishing of papers about the topic. A few markers of this change in emphasis are the substantial literature review work by Black and Wiliam³⁵, Sadler’s³⁶ work on formative assessment in instructional design, the work of Boud and Falchikov² on sustainable assessment in higher education or, in the context of medical education specifically, the work of Hodges and Lingard³⁷ on a reconsideration of competency and the need for a broader understanding of quality in assessment. Viewing assessment as a method for improvement and learning (i.e., the formative use of assessment information) instead of only for accountability (i.e., the summative use of assessment information), is one of the more significant changes and challenges in thinking about assessment.^{7, 17, 29, 31, 38} Taken together with the reconceptualisation of competence described above, CBME poses a significant assessment challenge and creates tensions when trying to design, interpret and implement assessment. For example, in a dual-purpose context, teachers are burdened with the responsibility not only to ensure learners have achieved acceptable standards, but also to create learning environments to support this process.

The assessment challenges associated with CBME and the reconceptualisation of these issues in the literature highlighted above gave rise to what can be called a ‘whole-system approach’ to assessment.

The rise of whole-system approaches to assessment

A whole-system assessment approach explicitly blends single assessments to achieve the formative and summative purposes of assessment for a variety of stakeholders, such as learners, teachers, and society-at-large.³¹ An example of a whole-system approach is coined by Van der Vleuten and Schuwirth¹⁸ as programmatic assessment, which theoretically aligns with the goals and curricula of CBME.^{39, 40} The model explicitly aims to improve high-stakes decision-making (i.e., assessment *of* learning) while simultaneously benefit and support student learning (i.e., assessment *for* learning). Instead of the problematised dichotomy between formative and summative assessment,^{20, 29, 41} assessment is conceptualised on a continuum of increasing stakes proportional to outcomes ranging from low-stakes assessments to high-stakes decisions. Such a continuum of increasing assessment stakes assumes every individual assessment combines the formative and summative purpose of assessment. Assessments are therefore dual-purpose: information generated with the assessment programme should be used to inform and support learning (i.e., the formative use of assessment information) and to inform decision-making and report on student learning (i.e., the summative use of aggregated assessment information). All kinds of assessments can be used as low-stakes assessments, ranging from multiple choice tests and direct observations of performance in the clinical workplace to self-assessments and peer feedback. They differ from high-stakes assessments in their consequences: low-stakes assessments should provide meaningful information for learners and teachers to support learning and hold little consequences for a learner based on their performance, however when aggregated, their purpose is to inform high-stakes decisions. Thus, the decision-moment is detached from the individual assessment moment. By moving away from an atomistic mastery-oriented approach towards a holistic approach to learning, the model aims to overcome the unintended consequences of competence-based assessment and the unjustifiable compromises required to make with single assessment instruments.^{18, 42}

In theory, programmatic assessment offers a solution to the assessment challenge of CBME, due to the opportunity to combine assessment purposes and different assessment formats, instruments and assessors, all used to inform learning and judgement and to construct a meaningful, holistic interpretation of competency development over time. The implementation of programmatic assessment in both undergraduate and post-graduate education is increasing and spreading around the globe rapidly. There is a high interest in this assessment model and more and more (medical education) institutions are implementing this approach in their curricula, especially for improving the alignment between what is considered competence and how this is assessed, the potential to improve progress decisions concerning such competence, and the potential to have assessment influence and inform learning in a more desirable way. Although appealing, implementing the approach is considered a major challenge.^{1, 7, 43, 44}

One of the bottlenecks of programmatic assessment in practice seems to be the use and the value of low-stakes assessment to the learner and teacher. The assumption is that lowering the stakes of individual assessments by detaching the decision moment from the assessment itself would 'take the sting out of assessment'^{22, 43} and more likely create learning opportunities by generating a continuous flow of information for learners that can be used to self-regulate their learning.^{38, 42} In this way, it could resolve the central problem with which I have started this introduction chapter. However, assessments designed or intended to be low-stakes may well be perceived as high-stakes by the learner and could potentially lead to a principle focus on assessment as a hurdle instead of a learning opportunity.^{43, 44, 46}

Learners' perceptions of assessment influence how they learn and teachers are said to play a powerful role in those perceptions.^{47, 48} Within a dual-purpose assessment system, teachers might experience significant dilemmas when navigating between their supportive roles in the learning process and their judgemental responsibilities as assessors of learners' performance.^{7, 49-51} As teaching happens through human interaction, the characteristics of teachers' interactions and relationships with learners can make a substantial difference to the kind of learning environment they create.⁵²⁻⁵⁴ Research outside the context of medical education has shown that learners' perceptions of teacher-learner relationships have a large impact on learners' motivation and effort, as well as on how they engage in various tasks.⁵⁵⁻⁵⁷ However, the literature and research approaches in medical education is still limited with respect to the effects of teacher-learner relationships on the use and uptake of assessment information.^{52, 53}

It does become clear how important the role and influence of learners' and teachers' reality is. Perceptions matter. The literature cited above highlights that the question is not if and how these different functions of assessment can be best combined, but rather how they, and the stakes involved, are perceived by those at the frontline of assessment and how such perceptions influence their actions.

The aim of this dissertation

Although interest in using assessment for more than decision-making and accountability might be higher than ever,^{7, 29, 31, 38} it entails major challenges too. This dissertation seeks to further explore, from a sociocultural and interpersonal perspective, the challenges related to simultaneously using assessment of learning and assessment for learning. Overall, I set out to gain an in-depth understanding of a whole-system assessment approach and the concept of assessment as a continuum of increasing stakes within such a system, and how these are understood by key stakeholders at the frontline of assessment practices: learners and teachers.

Key issues and questions that emerged from the literature described above are conceptualised as:

- How is low-stakes assessment constructed and perceived by learners and teachers? How do such conceptualisations influence learning and perceived learning opportunities with assessment? How do teachers and learners conceptualise their own role and responsibility in a whole-system approach?
- How do teacher-learner assessment relationships influence assessment perceptions? When and how do these assessment relationships afford (un)productive interactions and guidance for learning?
- What are the affordances of an assessment structure designed according to a programmatic approach in relation to its dual-purpose?

Methodological choices and dissertation overview

The research presented in this dissertation aims to explore and advance our understanding of assessment and its potential for learning within a whole-system assessment approach. The methodological choices in this work stem from qualitative research approaches which are particularly appropriate for such an explorative aim.^{58, 59} Although there are some differences and adaptations, I have used research methodologies originating from an interpretivist paradigm which hold similar stances and assumptions as those in which the questions and aim of this research are positioned. I embrace the idea and value of multiple realities and view myself, as well as the participants, as social beings who construct assessment according to their own values, beliefs and experiences. Consequently, I conceptualise assessment as a value-laden, socially and situated interpretative process. Continuing on this thought, I view the implementation of programmatic assessment as a structure, defined as the organised social institutions and patterns of interaction, or set of internally related objects or practices which can both enable and constrain individuals situated within such a structure.⁶⁰ I acknowledge the interdependence between individuals and this social structure: perceptions and actions are shaped by the perceived affordances of the structure as well as that the structure will shape the perceptions and actions of those same individuals. People can think and re-evaluate their situation and subsequently alter their course of action. Actions shape meanings and meanings evoke actions.⁶¹ Meanings are negotiated socially and historically, that is, they are not simply imprinted on individuals but are formed through interaction with others.⁶² These assumptions and beliefs reflect my socio-constructivist as well as pragmatic worldview, which strongly influenced all phases of this research project.

The dissertation gave me the opportunity to conduct a small series of studies in which each study could naturally build on one or more of the others, which allowed me to use developing insights to advance as I moved from one study to the next. I started my line of inquiry with an open and general research question in mind, which focussed on how low-stakes assessment is constructed and perceived by learners situated in a whole-system approach. This first study is presented in **Chapter 2**. I gathered data from five different programmes, in three different

countries, ranging from undergraduate to postgraduate medical education based on theoretical sampling. With semi-open interviews, I explored learners' experiences and interpretations of low-stakes assessment and how and why these were, or were not, beneficial for their learning and competency development. As I set out to explore *how* and *why* programmatic assessment and assessment stakes influence assessment perceptions, a constructivist grounded theory approach made an appropriate fit, which is also rooted in an interpretive tradition.^{58, 62, 63} Essentially, constructivist grounded theory is a direct methodological descendent of the pragmatist tradition and focusses on dynamic relationships between meaning and action.⁶¹ A constructivist grounded theory approach provides a way of showing and theorising how meaning and action influence each other. The anticipated product is an interpretive definition of theory in which theory emphasises interpretation and gives abstract understanding greater priority than explanation.⁶³ By means of a systematic and iterative process of co-constructing rich data, constant comparison, open and focused coding, memo-writing, and theoretical sampling, explained in more detail in the individual chapters, I have aimed to ensure quality and rigour in this interpretative and by nature subjective process and to gain an abstract understanding of the process involved. A constructivist approach allowed me to use my evolving and developing insights as an advantage. As such, the first study set the stage for the studies that followed and informed the research questions and the methodological considerations. Based on my understanding of learners' perceptions of assessment stakes and the identification of the powerful role teachers played in creating learning opportunities with the use of low-stakes assessment, I turned to the teachers' side of this story, a study presented in **Chapter 3**. In a similar manner, I explored teachers' conceptualisations of assessment in a whole-system approach and how these conceptualisations influenced teachers' perceived responsibility as assessors, how this influenced their role enactment and behaviours, and how they engaged with learners in assessment relationships.

The first two open studies were essential in gaining an abstract understanding of programmatic assessment in practice and as my understanding evolved, the initial open analytical process was more and more informed by my interpretation the data as well as by other theoretical frameworks. The role of using existing theories in this process is, and remains, controversial.⁶⁴ However, I believe that, especially in the context of conducting a series of studies, it is appropriate to consider how existing theoretical frameworks complement and extend my interpretation and understanding of the investigated process and that it is neither desirable nor possible to approach each new study with the same openness and naivety as I had when I first started my inquiry. This iterative and interpretative process of constant comparison is difficult to capture fully in the linear nature of this dissertation. Therefore, I will introduce two theoretical frameworks that influenced my interpretation and understanding of data in this introduction chapter, although they emerged after interaction with data. Taking a reflexive stance towards and during the whole research process to recognise the subjectivity I bring to this process as a researcher is therefore of utmost importance. In this spirit, the introduction chapter ends with a reflexivity paragraph.

Teacher-student assessment relationships

Interpersonal theory is pivotal to understand the influence of teacher-learner interactions and relationships on learners' assessment perception.⁶⁵ The theory states that to map perceptions of the self and of others in interpersonal communication, two independent dimensions are both sufficient and necessary. These dimensions are referred to as 'agency' and 'communion'.⁶⁵⁻⁶⁸ Agency 'connotes ideas of dominance, power, status, and control', and communion 'suggests love, affiliation, union, and friendliness'.⁶⁷ Applied to education,⁶⁹ teacher agency describes the level of teacher influence in the teacher-learner interaction and relationship and the dimension of communion refers to the level of warmth or friendliness a teacher communicates in interactions.^{67, 69-71} The interpersonal meaning of behaviour in interaction can be described as a combination of both dimensions.^{67, 70} In so-called interpersonal circumplex models, these two independent dimensions are combined in one model for describing the interpersonal meaning of behaviour in interactions or the position in the relationship.

Interpersonal theory has been used extensively to investigate teachers' and learners' interpersonal behaviours, interactions and relationships in secondary education,⁵⁶ and more recently in the supervision process within research projects in higher education and PhD projects.^{72, 73} In **Chapter 4** I take on this interpersonal theory-based perspective to investigate teacher-learner relationships in the context of assessment and the influence on learner's assessment perception. In this study, I conducted a focussed analysis using sensitising concepts from interpersonal theory to illuminate the influence of the teacher-learner relationship on learners' assessment perceptions. Given the hierarchical nature and location of power within any assessment relationship,^{74, 75} agency was of special interest and offered a relevant lens to further unravel and understand learners' assessment perceptions and engagement with assessment throughout this dissertation.

Agency

How agency is conceptualised, defined and what its significance is, has been the subject of numerous theoretical and philosophical debates and is therefore difficult to define in a conclusive or widely accepted manner.⁷⁶ Within this research project, I built on the interpretivist paradigm underpinning the development of assessment as a whole-system approach and conceptualise assessment as a socially and situated interpretative act. Within this paradigm, Ahearn⁷⁷ defines agency as "the sociocultural mediated capacity to act" which emphasises the situated nature of agency in sociocultural terms and the role of mediation. Mediation refers to how a structure or system, here the implementation of a whole-system approach, both enables and constrains the individual.⁶⁰ Mercer⁷⁶ argues that conceptualisations of agency need to highlight the multicomponent, intrapersonal nature of agency as well as the role of sociocultural-mediated processes. The emphasis is placed on the relationship between elements rather than the elements themselves. Aligned with this conceptualisation, the focus in this dissertation is on the relationship and interaction between agency and programmatic assessment as a social structure or system.

Furthermore, such a theoretical treatment of agency offers a relevant lens for the examination of interpersonal challenges in teacher-learner assessment relationships and the tension between independency and self-regulation on one hand and teachers' control and guidance on the other. Affordances, in this case of the whole-system approach, represent the interaction between contextual factors and learners' perceptions of them and the potential for learning inherent in this interaction.

Chapters 2 through **4** present a series of three empirical studies exploring assessment perceptions in a whole-system approach and the relationship with learning. The value and focus on the socially situated and contextual nature of assessment stimulated me to examine and review the state of research on programmatic assessment to enhance an understanding of the impact of the theory beyond the opportunities of my own empirical research. In **Chapter 7**, I present this knowledge synthesis. Using a systematic and integrative review study approach is considered appropriate to synthesise collective research in the context of health care professions education, which is complex and highly context dependent, because this type of review assists the understanding of the implication of collective research as a subjective process,⁷⁸ which is in line with the discourse and other methodological choices made in this dissertation.

Additionally, **Chapter 5** and **Chapter 6** offer two commentary papers. The first paper reflects on the role of and consequences for remediation in educational pathways when taking on a whole-system approach to assessment. **Chapter 6** offers a perspective on the decision-making process of assessment and the role of learner agency in this process when taking on a whole-system approach. In this chapter I juxtapose the way in which heuristics and other decision-making processes are commonly considered in the literature and argue to reconsider the role of learners in the high-stake decision-making process, especially if one aims to benefit learning with the use of assessment. Finally, the general discussion is presented in **Chapter 8** in which I aim to synthesise the results of the individual studies, answer the overall research question that initiated this dissertation, and contribute to the academic debate on the use of assessment for learning by taking a research-informed, but personally constructed stand on the matter.

Reflexivity

I recognise that data in this dissertation are co-constructed by interactions with the participants, as are the interpretations and meaning I gave to these data in collaboration with my research team. My own background and experiences, as well as those of my research team, shaped the questions we asked, which ones we pursued further, how we made sense of participants' responses, and how we drafted the manuscripts. To give the reader a sense of what and who is represented in these studies, it is important to share some relevant information about my own background. Like myself, readers of this dissertation most likely have some, if not a lot of, experiences and associations with assessment. For example, as a student being assessed or as a teacher assessing others. Assessment influenced my own educational trajectory and, at least partly, contributed to the position I now hold: the assessment of myself as undergraduate

student and the certificate which was awarded consequentially, allowed me to write, or maybe better phrased as 'to submit' my dissertation, as nobody would have prevented me from writing it in a private matter. Without it though, I would not have been allowed to access a graduate programme. I was fortunate to have mainly positive, challenging and sometimes even stimulating assessment experiences. So far, assessment never hold me back or prevented me from the opportunities I wanted to pursue. A researcher with more negative assessment experiences might have engaged with the concept of assessment differently. The professional roles I identify with most strongly during this research project are those of a teacher, an educational researcher, and an assessment consultant. I was initially trained and have worked as a teacher for several years. During this time, I was responsible for students' secondary school advise. I experienced first-hand how emotional such an event was for both the student involved as well as for their parents or caretakers. It influenced my initial rather sceptical attitude towards assessment and the role assessment plays in the educational pathways of students. I felt it distracted our attention and effort from the process I was much more interested and invested in, which was learning. As a teacher in higher and medical education, I started working with medical students and their teachers or supervisors concerning the topic of assessment. I have been privileged in working with highly motivated students and teachers, which makes teaching a very rewarding activity in my experience. I began to see assessment more as an entrée point for worthwhile discussions about the purpose and nature of education and I got more and more interested in the relationship between assessment and learning, more specifically in how assessment and assessment systems impact the people involved and how we could alter undesirable consequences of assessment. This dissertation allowed me to pursue this ambition on a conceptual and philosophical level while also influencing my teaching and assessment practices, and vice versa. My training as an educational scientist and the pursuit of this dissertation created more distance towards the topic as I gained more experiences with taking on different and unfamiliar perspectives on the matter. As a teacher and assessor, I felt much more emotionally involved in the process than I did in the role as researcher or as a consultant. Working as an assessment consultant, part of my responsibility concerned the quality assurance process of assessment. In this role, I mainly encounter problems and issues students and teachers have with assessment. Although I was aware that these experiences were not necessarily representative, in fact, they often are very selective and concern the exception rather than the rule, there is a risk in that these experiences coloured my thoughts and feelings about the relationship between assessment and learning. I struggled with the reality that in this role the focus of the majority of available resources were allocated to the minority of issues and problems, which fuelled my ambition to alter the impact of assessment to a more positive, worthwhile effort. These different roles, responsibilities and experiences with the different elements and sides of the assessment process all influenced the research project I engaged in. In that sense, it made me an insider. This influence is inevitable, and although this is neither good nor bad, it requires thoughtful consideration through reflexivity.⁶³ I have aimed to conduct this research in a critical and deliberate reflexive manner throughout the entire process. This was not done independently but in strong collaboration with a multidisciplinary research team with different backgrounds, expertise areas and involvement in programmatic assessment and education. When relevant, additional members were added

to the research team of an individual study, for example due to specific expertise. By means of the interplay between co-constructing rich data, constant comparison, coding, memo-writing and theoretical sampling I have developed conceptual categories and themes anchored in the data. During the coding process, memo-writing allowed me to continuously question everything and as such critically examine how the research was conducted and the judgements that were made throughout the research process. This process helped me to identify blind spots and avoid tunnel vision. Memo's included thoughts and discussions of the initial codes, the analytic and methodological questions that emerged during data collection and analyses, comparisons between fragments of data, emerging definitions of codes, their properties and characteristics and later on, examples and counterexamples of emerging codes and categories. I have aimed to make this process as transparent and credible as possible, which is represented in method sections of the individual studies and the reflexivity that runs throughout this dissertation. When possible, I have made data openly available and provided artifacts that were created, such as the interview protocols and codebooks, either integrated in the manuscript or in an appendix.

REFERENCES

1. 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):229-45.
2. Boud D, Falchikov N. *Rethinking assessment in higher education: learning for the longer term.* London: Routledge; 2007.
3. Laveault D, Allal L. *Assessment for Learning: Meeting the Challenge of Implementation.* Chambridge: Springer; 2016.
4. Gipps CV. *Beyond testing: towards a theory of educational assessment.* London: The Falmer Press; 1994.
5. Shepard LA. The Role of Assessment in a Learning Culture. *Educ Res.* 2016;29(7):4-14.
6. Cilliers FJ. Is assessment good for learning or learning good for assessment? A. Both? B. Neither? C. It depends? *Perspect Med Educ.* 2015;4(6):280-1.
7. Watling CJ, Ginsburg S. Assessment, feedback and the alchemy of learning. *Med Educ.* 2019;53(1):76-85.
8. Tavares W, Kuper A, Kulasegaram K, Whitehead C. The compatibility principle: on philosophies in the assessment of clinical competence. *Adv Health Sci Educ Theory Pract.* 2020;25(4):1003-18.
9. Govaerts M, Van der Vleuten CP. Validity in work-based assessment: expanding our horizons. *Med Educ.* 2013;47(12):1164-74.
10. Frenk J, Chen L, Bhutta ZA, Cohen J, Crisp N, Evans T, et al. Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. *Lancet.* 2010;376(9756):1923-58.
11. Frank J. *The CanMEDS 2005 physician competency framework: better standards - better physicians - better care.* 3rd printing, ed. Ottawa: Royal college of physicians and surgeons of Canada; 2006.
12. Swing SR. The ACGME outcome project: retrospective and prospective. *Med Teach.* 2007;29(7):648-54.
13. Ten Cate O. Competency-Based Postgraduate Medical Education: Past, Present and Future. *GMS J Med Educ.* 2017;34(5):Doc69.
14. Hodges B. Medical education and the maintenance of incompetence. *Med Teach.* 2006;28(8):690-6.
15. Govaerts MJB, Van der Vleuten CPM, Holmboe ES. Managing tensions in assessment: moving beyond either-or thinking. *Med Educ.* 2019;53(1):64-75.
16. Norcini J, Anderson B, Bollela V, Burch V, Costa MJ, Duvivier R, et al. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. *Med Teach.* 2011;33(3):206-14.
17. Schuwirth LWT, Van der Vleuten CPM. A history of assessment in medical education. *Adv Health Sci Educ Theory Pract.* 2020;25(5):1045-56.
18. Van der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. *Med Educ.* 2005;39(3):309-17.
19. Schuwirth L, Ash J. Assessing tomorrow's 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):555-9.
20. Eva KW, Bordage G, Campbell C, Galbraith R, Ginsburg S, Holmboe E, et al. Towards a program of assessment for health professionals: from training into practice. *Adv Health Sci Educ Theory Pract.* 2016;21(4):897-913.
21. Klass D. A performance-based conception of competence is changing the regulation of physicians' professional behavior. *Acad Med.* 2007;82(6):529-35.
22. Pugh D, Regehr G. Taking the sting out of assessment: is there a role for progress testing? *Med Educ.* 2016;50(7):721-9.

23. Mylopoulos M, Regehr G. Putting the expert together again. *Med Educ.* 2011;45(9):920-6.
24. Engestrom Y. *Expertise in Transition: Expansive Learning in Medical Work.* Cambridge: Cambridge University Press; 2018.
25. Bennett RE. Formative assessment: a critical review. *Assess Educ.* 2011;18(1):5-25.
26. Harlen W, James M. Assessment and Learning: differences and relationships between formative and summative assessment. *Assess Educ.* 2006;4(3):365-79.
27. Lau AMS. 'Formative good, summative bad?' – A review of the dichotomy in assessment literature. *J Furth High Educ.* 2015;40(4):509-25.
28. Pryor J, Crossouard B. A socio-cultural theorisation of formative assessment. *Oxf Rev Educ.* 2008;34(1):1-20.
29. Taras M. Assessment – Summative and Formative – Some Theoretical Reflections. *Br J Educ Stud.* 2005;53(4):466-78.
30. Scriven M. The methodology of evaluation. In: Tyler R, Gagne G, Scriven M, editors. *Perspectives on Curriculum Evaluation.* Curriculum Evaluation. Chicago: Rand, McNally and Co; 1967. p. 39-83
31. Norcini J, Anderson MB, Bollela V, Burch V, Costa MJ, Duvivier R, et al. 2018 Consensus framework for good assessment. *Med Teach.* 2018;40(11):1102-9.
32. Crooks TJ. The Impact of Classroom Evaluation Practices on Students. *Rev Educ Res.* 2016;58(4):438-81.
33. Bloom BS. Some theoretical issues relating to educational evaluation. In: Tyler RW, editor. *Educational evaluation: new roles, new means: the 63rd yearbook of the National Society for the Study of Education* Chicago: University of Chicago Press; 1969. p. 22-50.
34. Eva KW, Munoz J, Hanson MD, Walsh A, Wakefield J. Which factors, personal or external, most influence students' generation of learning goals? *Acad Med.* 2010;85:S102-5.
35. Black P, Wiliam D. Assessment and Classroom Learning. *Assess Educ.* 2006;5(1):7-74.
36. Sadler DR. Formative assessment and the design of instructional systems. *Instr Sci.* 1989;18(2):119-44.
37. Hodges B, Lingard L. *The question of competence: reconsidering medical education in the twenty-first century.* Ithaca: ILR Press; 2014.
38. Schuwirth LW, Van der Vleuten CP. Programmatic assessment: From assessment of learning to assessment for learning. *Med Teach.* 2011;33(6):478-85.
39. Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR. The role of assessment in competency-based medical education. *Med Teach.* 2010;32(8):676-82.
40. Norcini J, Holmboe E, Hawkins R. Evaluation challenges in the era of outcomes-based education. In: Holmboe E, Hawkins R, editors. *Practical Guide to the Evaluation of Clinical Competence.* 1e ed. Philadelphia, PA: Mosby/Elsevier; 2008. p. 1-9.
41. Taras M. Summative assessment: the missing link for formative assessment. *J Furth High Educ.* 2009;33(1): 57-69.
42. Van der Vleuten CP, Schuwirth LW, Driessen EW, Dijkstra J, Tigelaar D, Baartman LK, et al. A model for programmatic assessment fit for purpose. *Med Teach.* 2012;34(3):205-14.
43. Bok HG, Teunissen PW, Favier RP, Rietbroek NJ, Theyse LF, Brommer H, et al. Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Med Educ.* 2013;13:123.
44. 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):487-98.
45. Harrison C, Wass V. The challenge of changing to an assessment for learning culture. *Med Educ.* 2016;50(7):704-6.

CHAPTER 1

46. 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):276-84.
47. Ames C. Classrooms: Goals, structures, and student motivation. *J Educ Psychol.* 1992;84(3):261-71.
48. Segers M, Nijhuis J, Gijsselaers W. Redesigning a learning and assessment environment: The influence on students' perceptions of assessment demands and their learning strategies. *Stud Educ Eval.* 2006;32(3):223-42.
49. 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):442-67.
50. Rea-Dickins P. Understanding teachers as agents of assessment. *Language Testing.* 2004;21(3):249-58.
51. William D. Embedded formative assessment. Bloomington, IN: Solution Tree Press; 2011.
52. Haidet P, Stein HF. The Role of the Student-Teacher Relationship in the Formation of Physicians : The Hidden Curriculum as Process. *J Gen Intern Med.* 2006;21(S1):S16-S20.
53. Telio S, Ajjawi R, Regehr G. The "educational alliance" as a framework for reconceptualizing feedback in medical education. *Acad med.* 2015;90(5):609-14.
54. Ramani S, Post SE, Konings K, Mann K, Katz JT, Van der Vleuten C. "It's Just Not the Culture": A Qualitative Study Exploring Residents' Perceptions of the Impact of Institutional Culture on Feedback. *Teach Learn Med.* 2017;29(2):153-61.
55. Den Brok P, Brekelmans M, Wubbels T. Interpersonal Teacher Behaviour and Student Outcomes. *Sch Eff Sch Improv.* 2004;15(3-4):407-42.
56. Wubbels T, Brekelmans M. Two decades of research on teacher–student relationships in class. *Int J Educ Res.* 2005;43(1-2):6-24.
57. Wentzel KR. Motivating Students to Behave in Socially Competent Ways. *Theory Pract.* 2003;42(4):319-26.
58. Corbin J, Strauss A. *Basics of Qualitative Research (3rd ed.): Techniques and Procedures for Developing Grounded Theory.* California: Thousand Oaks; 2008.
59. Creswell JW. *Educational Research: Planning, Conducting and Evaluating Quantitative and Qualitative Research.* Essex: Pearson Education Limited; 2014.
60. Pearson J. Assessment of agency or assessment for agency? A critical realist action research study into the impact of a processfolio assessment within UK HE preparatory courses for international students. *Educ Action Res.* 2020;29(2):259-75.
61. Charmaz K. The Power of Constructivist Grounded Theory for Critical Inquiry. *Qual Inq.* 2016;23(1):34-45.
62. Creswell JW. *Qualitative inquiry and research design: choosing among five approaches.* 3rd ed. London: Sage Publications; 2013.
63. Charmaz K. *Constructing Grounded Theory. A Practical Guide Through Qualitative Analysis.* Thousand Oaks: Sage Publications, Inc.; 2006.
64. Watling CJ, Lingard L. Grounded theory in medical education research: AMEE Guide No. 70. *Med Teach.* 2012;34(10):850-61.
65. Horowitz LM. *Interpersonal Foundations of Psychopathology.* Washington, DC: American Psychological Association; 2004.
66. Bakan D. *The duality of Human Existence: Isolation and Communication in Western Man.* Boston, MA: Beacon Press; 1966.
67. Gurtman MB. Exploring Personality with the Interpersonal Circumplex. *Soc Personal Psychol Compass.* 2009;3(4):601-19.

68. Fournier MA, David DSM, Zuroff DC. Origins and Applications of the Interpersonal Circumplex. In: Horowitz LM, Strack S, editors. *Handbook of Interpersonal Psychology: Theory, Research, Assessment, and Therapeutic Interventions*. Hoboken, New Jersey: John Wiley & Sons, Inc.; 2011. p. 57-73.
69. Wubbels T, Brekelmans M, Mainhard T, Den Brok P, Van Tartwijk J. Teacher–student relationships and student achievement. In: Wentzel KR, Ramani GB, editors. *Handbook of Social Influences in School Contexts: Social-Emotional, Motivation, and Cognitive Outcomes* New York: Routledge; 2016. p. 127-42
70. Wiggins JS. *The five-factor model of personality : theoretical perspectives*. New York: Guilford Press; 1996.
71. Pennings HJM, Brekelmans M, Claessens LCA, Van Tartwijk J, Sadler P, Van der Want AC. Interpersonal adaptation in teacher-student interaction. *Learn Instr*. 2018;55:41-57.
72. Mainhard T, Van der Rijst R, Van Tartwijk J, Wubbels T. A model for the supervisor–doctoral student relationship. *High Educ*. 2009;58(3):359-73.
73. De Kleijn RAM, Meijer PC, Pilot A, Brekelmans M. The relation between feedback perceptions and the supervisor–student relationship in master's thesis projects. *Teach High Educ*. 2013;19(4):336-49.
74. Heron J. Assessment revisited. In: Boud D, editor. *Developing Student Autonomy in Learning*. Second ed. London: Kogan Page Limited; 1981. p. 55-68.
75. Reynolds M, Trehan K. Assessment: A critical perspective. *High Educ Stud*. 2000;25(3):267-78.
76. Mercer S. The complexity of learner agency. *Int J Appl Lang Stud*. 2012;6(2):41-59.
77. Ahearn LM. Language and Agency. *Annu Rev Anthropol*. 2001;30:109-526.
78. Bearman M, Dawson P. Qualitative synthesis and systematic review in health professions education. *Med Educ*. 2013;47(3):252-60.



CHAPTER 2

Stakes in the eye of the beholder: An International Study of Learners' Perceptions within Programmatic Assessment

Published as:

Schut, S., Driessen, E., Van Tartwijk, J., Van der Vleuten, C., & Heeneman, S. (2018). Stakes in the eye of the beholder: an international study of learners' perceptions within programmatic assessment. *Medical Education*, 52 (6), 654-663. <https://doi.org/10.1111/medu.13532>

ABSTRACT

OBJECTIVES Within programmatic assessment, the ambition is to simultaneously optimise the feedback and the decision-making function of assessment. In this approach, individual assessments are intended to be low stakes. In practice, however, learners often perceive assessments designed to be low stakes as high stakes. In this study, we explored how learners perceive assessment stakes within programmatic assessment and which factors influence these perceptions.

METHODS Twenty-six learners were interviewed from three different countries and five different programmes, ranging from undergraduate to postgraduate medical education. The interviews explored learners' experience with and perception of assessment stakes. An open and qualitative approach to data gathering and analyses inspired by the constructivist grounded theory approach was used to analyse the data and reveal underlying mechanisms influencing learners' perceptions.

RESULTS Learners' sense of control emerged from the analysis as key for understanding learners' perception of assessment stakes. Several design factors of the assessment programme provided or hindered learners' opportunities to exercise control over the assessment experience, mainly the opportunities to influence assessment outcomes, to collect evidence and to improve. Teacher-learner relationships that were characterised by learners' autonomy and in which learners feel safe were important for learners' believed ability to exercise control and to use assessment to support their learning.

CONCLUSIONS Knowledge of the factors that influence the perception of assessment stakes can help design effective assessment programmes in which assessment supports learning. Learners' opportunities for agency, a supportive programme structure and the role of the teacher are particularly powerful mechanisms to stimulate the learning value of programmatic assessment.

INTRODUCTION

Programmatic assessment as a new approach to assessment is emerging rapidly within medical education.¹⁻³ This approach is used in various medical school programmes around the world, ranging from undergraduate to postgraduate.⁴⁻⁶ Programmatic assessment can be used as a framework when designing assessment programmes that are aimed at optimising both the learning and the decision-making function of assessment.⁷ A growing body of evidence to support the value of programmatic assessment is emerging, and although research shows the first positive results that this assessment approach might be beneficial for supporting the development of self-regulated learning,^{5, 8, 9} implementing this approach is a challenge and many of the principles are still uncertain in practice.^{4, 5, 10} There is an urgent need for empirical verification of the principles and concepts underlying the theoretical model of programmatic assessment.

One of the important concepts within programmatic assessment is that assessment is proposed as a continuum with a proportional relationship between what is at stake and the number of individual assessments.⁷ Each individual assessment itself has limited consequences for the learner (i.e., is low stakes) but the consequences of the evaluation of the aggregated assessments can be substantial when they are used for a decision about, for instance, graduation or promotion (i.e., high stakes). Lowering the stakes of the individual assessment is supposed to optimise and benefit the learning potential of programmatic assessment, and provide learners with a continuous flow of information about their performance.¹¹ However, researchers have reported a mismatch between the designers' intentions to develop low-stakes assessments to stimulate and optimise learning, and learners' perceptions of these assessments as high stakes and summative.^{4, 5} This potentially leads learners to focus on each individual assessment as a hurdle and not as a learning opportunity.^{12, 13} Furthermore, it raises the question of whether the meaning of assessment stakes as defined in the theoretical model of programmatic assessment (the consequences following an assessment) is the same for learners whose performance is being assessed.

The impact of any assessment system on learning is mediated by learner's perceptions.^{13, 14} Insight into these perceptions is therefore crucial for understanding what low and high stakes mean to the learner, and how and why assessment enables, or fails to enable, learners to optimise and self-regulate their learning using these assessments. Therefore, the current study aims to gain more insight into how assessment stakes are perceived by learners and which factors influence learners' perceptions.

METHODS

Sample

We purposively selected different assessment programmes and interviewed learners from multiple institutes, countries and educational phases about their assessment experiences within programmatic assessment. We used an open and qualitative approach to data gathering and analyses, inspired by constructivist grounded theory.^{15, 16} The inclusion criteria were: (i) a programmatic approach to assessment is used, including low-stakes assessments aiming to provide learners with information about their progress, and high-stakes decisions regarding learners' progress are based on the evaluation of the aggregation of multiple low-stakes assessments; and (ii) there is a stable implementation of programmatic assessment over a longer period to minimise interference in the perceptions of assessment stakes due to suboptimal implementation issues. It was expected that learners' experiences with and views on assessment and the stakes involved could vary with their level of training and the process of enculturation into different learning communities. Therefore, programmatic assessment practices from pre-clinical and clinical phases were purposively selected: from pre-clinical undergraduate education (setting A); from clinical undergraduate education (setting B); and from clinical postgraduate medical education (setting C). In all selected programmes there is diversity in assessment formats including a portfolio. In these portfolios, learners collect, combine and reflect on the assessment information with the aim of self-regulating their learning, supported by a mentor. The structure and characteristics of the different assessment programmes are presented in Table 1. Within these different programmes, participants were selected based on their assessment experience: all participants have experienced at least one full feedback loop (i.e., the process of multiple low-stakes assessments and at least one high-stakes decision).

Data collection

E-mails inviting learners to participate in one-to-one interviews were sent by local faculty members to all selected participants. A convenience sampling approach was taken based on learners' availability at predetermined times. A total of 26 respondents participated in individual, semi-structured interviews. Open-ended questions were posed by one interviewer (SS), who asked participants to describe their assessment experiences, including if and why they considered the assessment meaningful for their learning and what they perceived the stakes to be, and to reflect on the consequences that followed based on their performance. When the participant did not mention high-stakes assessments, the interviewer asked him or her to reflect on an assessment with consequences designed to be high stakes, for example a certification examination or progress decisions based on a portfolio, and if or how this was different, in order to fully understand the participant's assessment experience. All sessions were recorded and transcribed verbatim. Interviews and analyses were conducted iteratively, allowing early insights, conceptual ideas and unexpected findings to shape subsequent data collection.¹⁵ Data were collected between April 2016 and November 2016. Participants received a small compensation (a \$10 gift card). Ethical approval was obtained from the Dutch

Association for Medical Education Ethical Review Board (NVMO-ERB668 on 1 March 2016), the Dalhousie Health Sciences Research Ethics Board (REB#2016-3882 on 25 July 2016) and the Cleveland Clinic Institutional Review Board (IRB#16-1261 on 21 September 2016).

Data analysis

Interview data were analysed using a constant comparative approach.¹⁵ Independent analysis of the first four transcripts using an open coding strategy was carried out by SS and SH. During this process, coding results and relations between codes were discussed constantly. Differences were discussed until consensus was reached. This process resulted in initial codes and preliminary themes, which were used by the first author (SS) for coding of the next four transcripts. When new codes and themes emerged, these transcripts too were independently analysed by the second researcher (SH) to test the fit and relevance of the new codes and themes. Necessary adaptations to the interview questions were made for the subsequent interviews. Through coding and constant comparison, data were organised around two main categories: programme factors and (inter) personal factors. Several discussions with all members of the research team were organised in order to reach consensus on the themes that emerged, on the depth of the preliminary analysis and on the relationships between codes and categories in order to raise the analytical level from categorical to conceptional. Furthermore, two members of the research team (ED and JvT) read two additional transcripts to review the data and to ensure a fit with the codes and discussed themes. Data collection and analysis continued until theoretical sufficiency was reached, defined as 'the stage at which categories seem to cope adequately with new data without requiring continued extensions and modifications'.¹⁷ Theoretical sufficiency was proposed by Dey¹⁷ and offers a more nuanced alternative to saturation to deal with issues concerning the sense of completeness and certainty implied by theoretical saturation.¹⁸ The following criteria were used: (i) new data could be fitted in categories that were already developed; (ii) no new insights, themes, issues or counter-examples or cases arose; and (iii) consensus within the research team was reached about the notion of sufficiency with the collected and analysed data.¹⁵⁻¹⁷ All interviews were then re-read by the first researcher to ensure that no relevant information was missed.

Reflexivity

We acknowledge that data in this study are co-constructed by interactions with the participants, as are the interpretations and meaning we gave to these data.¹⁵ To prevent biases as much as possible, we brought together a multidisciplinary research team: SS and ED have a background in educational sciences, CvdV in psychology, JvT in sociology and SH in biomedical sciences. SS, ED, CvdV and SH all have involvement in programmatic assessment in medical education. To avoid tunnel vision in our interpretation of the data, we brought in an outsider perspective: JvT works in the social sciences and in teacher education and is not directly involved in medical education.

CHAPTER 2

TABLE 1. Summary of characteristics of the selected assessment programmes

	Preclinical (n=11)			Clinical (n=15)	
	Undergraduate medical education			Post-graduate medical education	
	A1	A2	B1	C1	C2
Institute	Cleveland Clinical Lerner College of Medicine, Ohio USA	Faculty of Health, Medicine and Life Science, Maastricht University	Faculty of Health, Medicine and Life Science, Maastricht University	Dalhousie University Department of Family Medicine, Canada	Maastricht University Medical Centre, Maastricht
Programme	5-year graduate-entry programme, Physician Investigator	4-year graduate-entry master programme Physician-Clinical Investigator	6-year bachelor-master programme Medicine	2-year family Medicine residency programme	3-year family Medicine residency programme
Course/ phase	Year 1 and 2	Year 2	Year 2 of the master phase, the 12 weeks clinical rotation Family Medicine (last of five clinical rotations in the master phase)	Year 2 of the residency programme	End year 1 or year 3 of the residency programme
Interviews*	17(M), 18(M), 19(M), 20(F), 21(F), 22(M)	1(F), 2(F), 3(M), 4(F), 9(F)	5(F), 6(F), 7(F), 8(F)	11(F), 12(F), 13(F), 14(M), 15(F), 16(F)	10(F), 23(M), 24(M), 25(F), 26(F)
Learners p/y	32	50	330	15	17
Low-stake assessments**	Weekly SAQ's & CAPPs, PBL (peer) evaluations, direct observations, OSCE's, Journal Club, periodic reviews	Knowledge (in- and end-of-block) tests, progress tests, OSCEs, direct observations, scholarly projects, variety of assignments	Workplace-based performance evaluation forms (Mini-CEX's, field-notes), formative knowledge test, progress tests, case-based discussions, variety of assignments	Evaluation objectives, field notes, reflective discussions, narrative, OSCEs, presentations, scholarly project, ITAR, periodic reviews	Formative evaluations (knowledge & skills), National knowledge progress test Family Medicine, scholarly project, presentation, self-evaluation, video assessments

* M: male; F: Female

**SAQ's: Self-Assessment Questions, CAPPs: open book Concept Appraisals, PBL: Problem based learning tutorials, OSCE: Objectives Structured Clinical Examinations, Mini-CEX: Mini clinical evaluation exercise, ITAR: narrative in-training assessment reports

RESULTS

Overall, learners shared the same definition of stakes as defined within the model of programmatic assessment; that is, the consequences following an assessment. However, these consequences were not primarily considered as the proposed continuum in the programmatic assessment model, but rather as a dichotomy. Assessment comes with stakes (i.e., with consequences) or no stakes at all (i.e., no consequences); *'It doesn't count, nobody cares, it's not like you have to remediate or take a resit or whatever'* (A2). Assessment as a continuum of stakes was recognised but appeared much more complex, encompassing more than just consequences following an assessment. In all different programmes, learners' conceptualisation of assessment stakes as a continuum was strongly related to their perceived ability to act, control and make choices within the learning and assessment environment. Several design factors of the assessment programme influenced learners' opportunities to exercise control. Whether or not learners acted upon these opportunities depended on the interplay between experience and confidence, as well as the relationship with others in the assessment environment such as teachers. The results are presented as (i) the opportunities for learners' control within the assessment programme and (ii) factors influencing learners' believed ability to exercise control.

Opportunities for control within the assessment programme

Several programme design factors influenced learners' opportunities to exercise control and with that the perception of stakes. These aspects are described below, with participants' quotes to illustrate the themes.

Opportunities to influence outcomes

What was being assessed and which format was used for the assessment influenced the perceived stakes. In the case of making progress in generic competencies (e.g., communication, collaboration and professionalism), learners experienced multiple perspectives on the requirements of these competencies, often without a clear standard or norm, resulting in a perception of more influence on the required outcomes. By contrast, most learners considered standardised knowledge tests as high stakes and associated these assessment tasks with a fixed norm to be achieved. They experienced success in such assessments as being able to 'find the correct answer', according to a pre-constructed test and answer key, which led to a feeling of being highly dependent on the content, quality and relevance of the specific test. This caused a perception of little to no control over the assessment and outcomes, especially when this type of assessment resulted in grades:

You either know it or you don't when it comes to knowledge. Whereas I guess when you're talking to your preceptor about a field note, it's less measurable outcomes. So, it's more so about reflecting and just talking through something, it's more fluid than grades. (C1)

Furthermore, the opportunity to interact with the assessor (e.g., during an oral examination, or when an assessor would interact with the learner during direct observation) was perceived as a potential influence on the assessment outcome. Learners indicated that interaction with the assessor provided more opportunities to show their progress and abilities and made them feel more in control over the process and the outcome of the assessment. This lowered the perceived stakes. However, interaction with the assessor could also raise the perceived stakes: learners thought this carried the risk of losing face, especially when the assessor was intimidating, an important role model or worked in a discipline of interest.

Opportunities to collect evidence

In all programmes, learners collected evidence within a portfolio, with the aim of monitoring and showing their progress. However, programmes varied in the freedom learners had to collect and select evidence. Some programmes gave learners the opportunity to initiate an assessment, for example by allowing learners to assess their knowledge development by taking formative self-tests or encouraging learners to ask for direct observation on the learners' own terms. This feeling of control not only lowered the perceived stakes, but more importantly also seemed to make the assessment feel more relevant:

You have more control over the assessment [initiating a Mini-CEX] and then you can focus the assessment to what is important to you. You can tailor it to what you need at that moment. That makes it low-stake and more meaningful. (C2)

The perceived stakes were lower when results or follow-ups were not automatically accessible to others and learners could control what was shared. Learners experienced more choice and felt more in control when given the opportunity to select their own evidence for the portfolio:

To me those [own evidence] would be the lowest stakes, because you are not expected to collect them, they are not expected to be in there [the portfolio]. (A1)

Opportunities to improve

The procedures offered by programmes to improve earlier insufficient performance influenced learners' perceived control of the impact of each individual assessment. An important factor was whether or not opportunities for improvement were integrated into the educational programme. When this had to be done next to the regular curriculum or assessment activities, the time investment needed for improvement felt like an overload, the stakes became higher, and learners were more motivated to avoid this:

It [curriculum] is already overloaded [...] For me that is also the incentive to just want to pass and get it over with, I think otherwise the other things will be in jeopardy. (A2)

Most programmes provided multiple complementary assessments that were meant to facilitate more opportunities for learners to show progress and improvement. This lowered the stakes of the individual assessment. Not being solely dependent on one individual ‘snapshot’ gave learners more feeling of being in control, because of multiple opportunities to show and improve on their performance, especially when the focus was on trends or reoccurring feedback messages:

I have lots of different people talking about my professionalism and so, each additional one has less impact. It's okay to get some negative feedback, because you have a lot. So, you have some negative and some positive, just, you know that's how it tends to balance out. (B1)

Although the number of complementary assessments influenced the perception of assessment stakes positively, this also came with reaching a point of so-called ‘overkill’, in which the assessment became meaningless and a checkbox activity to meet the requirements of the programme:

It almost becomes a hunt on evaluations. And it's not about the quality or their usefulness anymore, but just about the quantity. (B1)

Nonetheless, learners did not always recognise the coherence or complementary nature of multiple assessments, causing them to perceive the assessment as isolated and therefore high stakes. When grades were used for individual assessments, and learners could correct insufficient performance by ways of averaging multiple results, this contributed to learners’ understanding of the coherence. However, receiving grades also contributed to competition amongst learners, anxiety and a performance orientation, which raised the stakes. The following is an example of a learner reflecting on his transition from a previous assessment environment with grades to his current environment without grades:

In undergraduate studies a test would make me nervous and anxious and worried about how well I was going to do. But because these assessments don't have the same consequences [receiving grades], because they are just to help me identify what to study, I don't feel the same nervousness that I did before. It's the good without the bad. (A1)

Factors influencing learners’ believed ability to exercise control

Learners used opportunities provided by the assessment programme for control, when they believed they had the ability to exercise control. This belief was influenced by personal attributes, as well as the relationship experienced with teachers. In effect, the factors influencing learners’ believed ability to exercise control are presented as (i) the interplay between learners’ experience and confidence and (ii) the influence of teachers.

The interplay between experience and confidence

Previous experience of assessments influenced the perceived stakes within all programmes. Most learners were accustomed to defining success as being top of their class and getting high scores or grades:

Assessments, they were always a bit stressful for me. [...] There was always something connected to them, like proceeding to the next year for example. And also, like with my parents and grandmothers, if I got a good grade, then there is a certain reward. And I think that is also like a bit of conditioning. (C2)

In programmes previously attended, this was often rewarded and even viewed as a necessity. An example is the situation when admission to a medical school required a high secondary school grade point average. Assessment was then associated with pressure for high performance, insecurities and fear of failing. Such assessment experiences had a strong impact. New experiences were required before these associations were replaced with a more learning-oriented perception of assessment. Learners had to gain confidence in the meaning and consequences of the low-stakes assessment, which contributed significantly to the perception of stakes. First-time experiences with low-stakes assessments were unanimously perceived as high stakes, especially when learners did not fully understand what was expected, or what could happen when they were unable to meet the demands: *'I think a lot of the anxiety was caused by us not knowing exactly what was going to happen [if we would perform poorly on an OSCE]'* (A1). The more familiar learners became with such assessments, the less anxious they felt.

The influence of teachers

The believed ability to exercise control and therefore the perception of assessment as low stakes seemed strongly dependent on learners' relationship with their teachers. When learners felt the teacher was their advocate, facilitated learning, and allowed them to experiment and to take control, they felt safe and able to interpret low-stakes assessments as low stakes and meaningful for learning. The assessment environment was then described as a safe place to learn and experiment: *'I feel very comfortable looking stupid'* (A1) and *'I think it's hard to feel the fear of failing'* (C1). However, some learners felt that the relationship was characterised by an unequal power balance that influenced their perception of assessment stakes:

So, he [the teacher] has all the power. That's how it feels to me. He has a lot to say about it. The things he considers important, he picks them out and focuses on them. And the consequence they have, I think are much bigger than the consequences such a test should actually have. (B1)

For learners to take control, the teacher needed to provide them with the opportunity to exercise control.

DISCUSSION

The theoretical assumption underlying the proposed continuum of assessment stakes within programmatic assessment, is that low-stakes assessments create learning opportunities and generate a continuous flow of information for learners that can be used to self-regulate their learning.^{7, 11} This requires assessment that is intended or designed to be low stakes, to be perceived as such by the learner.¹²⁻¹⁴ This study identified the feeling of being in control to be essential for learners' perception of assessment stakes, and identified factors that allowed or hindered learners' opportunity to exercise control. This is strongly linked to the concept of agency, referred to as learners' perceived ability to act, control and make choices within the learning and assessment environment.^{12, 13, 19} The value and importance of learners' agency for continuous development using assessment has already been highlighted by others^{1, 8, 12, 19-21} and affects the ability or willingness to learn from assessment.^{12, 13, 22} What this study contributes is insight into how learners' agency is negotiated in the context of programmatic assessment and what enables and constrains its emergence.

Different programme factors provided or hindered learners' opportunities to take control over the assessment experience. Standardised assessments provide little opportunity for learners' agency. Although necessary and understandable, standardisation places the control at the programme level, leaving little space for the individual learner to exercise control. This might even alienate learners from their learning and assessment experience.^{23, 24} A sense of agency was, however, encouraged when the programme allowed learners to initiate their own assessment, and when learners were enabled to select evidence of progress. This has the potential to engage learners more actively in the assessment process.^{12, 19, 25}

Increasing the number of opportunities for learners to monitor and show progress, even with standardised knowledge tests, can be another strategy to lower the stakes. Programme designers should take care, however, not to create an assessment overload for both learners and faculty members. In addition, the link between individual low-stakes assessments must be clear for the learner in order to provide direction on how low-stakes assessment can and should be used to support learning.²⁶ Although using grades and opportunities to compensate can highlight the coherence amongst complementary assessments, this may have the adverse effect of encouraging different, less desirable, study strategies and behaviours.²⁷ Providing grades has the implicit risk of encouraging a focus on outcomes and competition rather than stimulating a focus on continuous improvement.^{12, 28} Not providing grades seemed to enable a shift to a learning orientation, described as one in which the learner's goal is to improve.^{29, 30} Linking complementary low-stakes assessments was better accomplished in programmes that highlighted the influence of low-stakes assessments on learners' improvement plans. The focus should be on using information generated by low-stakes assessment to analyse and reflect upon learning progress and how this should direct future learning. We could consider giving more opportunity for the learner to control the appropriate number of assessments needed to show progress and improvement, rather than setting up quantity requirements. This can make the assessment

CHAPTER 2

experience a more personal inquiry and create ownership over the plan of improvement. Thinking this way does not take away the need for some type of consequences: when learners perceived little incentive to address their weaknesses or to act upon information concerning their strengths and weaknesses, low-stakes assessment rarely led to a focus on improvement. The implementation of the so-called post-assessment process (the follow-up activities or reflective tasks) is therefore essential,³¹ both in the design of a supportive programme structure (i.e., facilitating room for improvement) and in the role of the teacher (i.e., valuing and stimulating improvement versus performance).

Last, learners' confidence and their believed ability to exercise control seemed to increase over time. Novices within programmatic assessment needed time to adjust to and get familiar with the new assessment approach. Associations and experiences with high-stakes assessment need to be phased out and teachers need to adjust the level of guidance and direction to the experience level of learners.³² Moreover, when the teacher–learner relationship can be characterised as safe and with autonomy for the learner, learners are more likely to use assessment to support their learning. Within programmatic assessment, learners should be allowed independence and control over the assessment process. Only then, will learners perceive assessment as low stakes.

Our results fit well with the calls to create a shared responsibility between learners and teachers within the assessment process^{1, 33} and with the need to create a learning environment where dialogue can flourish to engage learners actively with feedback and assessment.^{21, 34} This challenges teachers to reconcile their responsibility for stimulating and evaluating development.^{1, 12, 20} Teachers are fundamental for creating this safe learning environment and utilising the potential of programmatic assessment.

Limitations

This study has several limitations. Firstly, learners within medical education are typically characterised as high achievers, selected through rigorous course admission procedures. Many learners in this study referred to themselves as such and only a few of them had experienced failure in relation to assessment or progress decisions. The perception of assessment stakes might work differently for low-achieving learners. Future work could explore the relevance of our concepts and mechanisms in other (academic) settings by including programmes outside medical education or purposefully selecting low-achieving learners.

Secondly, participants received a small compensation for their time and effort. All ethical committees involved approved this. Without exceptions, participants seemed motivated to further our understanding of their assessment experience and to contribute to the quality of the educational programme, but of course we can't exclude that the small reward may have biased our sample.

Thirdly, the number of interviews per programme was limited. However, this study was not designed to compare different implementations of programmatic assessment. We included a

range from undergraduate to postgraduate medical education to understand the underlying mechanisms that appeared in different programmes and to identify influencing design factors. Although the decision to sample different phases of medical education, and staying within one discipline, was made on methodological grounds, this could have influenced the representation of certain themes. Differences between programmes, class sizes and disciplines may also impact the implications of our findings and recommendations for other practices. We therefore stress the importance of replicating our study in different contexts and learning cultures for further understanding and transferability.

Finally, given the importance of the role of the teacher, future studies should triangulate students' self-reported perceptions by exploring the perceptions of teachers regarding the stakes of assessment.

CONCLUSION

This study identified factors that influenced the stakes learners perceived within a programmatic approach to assessment. Learners perceive assessments more as low stakes when we provide opportunities for learners to exercise control and create an assessment environment that embraces possibilities for improvement. Environments wherein learners can freely discuss their weaknesses and uncertainties, and teachers encourage learners to share learning needs, invest in their improvement plans, and provide guidance and direction needed for achieving progress and improvement. In summary, learners' opportunities for agency, a supportive structure and the role of the teacher are particularly powerful factors in stimulating the perception of assessment as low stakes and enhancing the learning potential of assessment. Knowledge and understanding of the identified factors can help educational developers to design effective programmes of assessment that could increase the learning value of assessment.

REFERENCES

1. Eva KW, Bordage G, Campbell C, Galbraith R, Ginsburg S, Holmboe E, et al. Towards a program of assessment for health professionals: from training into practice. *Adv Health Sci Educ Theory Pract.* 2016;21(4):897-913.
2. Van der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. *Med Educ.* 2005;39(3):309-17.
3. Schuwirth L, Van der Vleuten C, Durning SJ. What programmatic assessment in medical education can learn from healthcare. *Perspect Med Educ.* 2017;6(4):211-5.
4. Bok HG, Teunissen PW, Favier RP, Rietbroek NJ, Theyse LF, Brommer H, et al. Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Med Educ.* 2013;13:123.
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):487-98.
6. 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.
7. Van der Vleuten CP, Schuwirth LW, Driessen EW, Dijkstra J, Tigelaar D, Baartman LK, et al. A model for programmatic assessment fit for purpose. *Med Teach.* 2012;34(3):205-14.
8. Altahawi F, Sisk B, Poloskey S, Hicks C, Dannefer EF. Student perspectives on assessment: experience in a competency-based portfolio system. *Med Teach.* 2012;34(3):221-5.
9. Sundre DL, Kitsantant A. An exploration of the psychology of the examinee: Can examinee self-regulation and test-taking motivation predict consequential and non-consequential test performance? *Contemp Educat Psychol.* 2004;29(1):6-26.
10. Harrison CJ, Konings KD, Molyneux A, Schuwirth LW, Wass V, Van der Vleuten CP. Web-based feedback after summative assessment: how do students engage? *Med Educ.* 2013;47(7):734-44.
11. Schuwirth LW, Van der Vleuten CP. Programmatic assessment: From assessment of learning to assessment for learning. *Med Teach.* 2011;33(6):478-85.
12. 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):276-84.
13. Cilliers EJ, 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.
14. Segers M, Nijhuis J, Gijsselaers W. Redesigning a learning and assessment environment: The influence on students' perceptions of assessment demands and their learning strategies. *Stud Educ Eval.* 2006;32(3):223-42.
15. Watling CJ, Lingard L. Grounded theory in medical education research: AMEE Guide No. 70. *Med Teach.* 2012;34(10):850-61.
16. Corbin J, Strauss A. *Basics of Qualitative Research (3rd ed.): Techniques and Procedures for Developing Grounded Theory.* California: Thousand Oaks; 2008.
17. Dey I. *Grounding grounded theory : guidelines for qualitative inquiry.* San Diego: Academic Press; 1999.
18. Morse JM. The Significance of Saturation. *Qual Health Res.* 2016;5(2):147-9.
19. Cilliers EJ, 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.

20. Watling C. The uneasy alliance of assessment and feedback. *Perspect Med Educ*. 2016;5(5):262-4.
21. Boud D, Molloy E. Rethinking models of feedback for learning: the challenge of design. *Assess Eval High Educ*. 2013;38(6):698-712.
22. Wiliam D. *Embedded formative assessment*. Bloomington, IN: Solution Tree Press; 2011.
23. Mann SJ. Alternative Perspectives on the Student Experience: Alienation and engagement. *High Educ Stud*. 2010;26(1):7-19.
24. Kahu ER. Framing student engagement in higher education. *High Educ Stud*. 2013;38(5):758-73.
25. Shepard LA. The Role of Assessment in a Learning Culture. *Educ Res*. 2016;29(7):4-14.
26. Olupeliyawa A, Balasooriya C. The impact of programmatic assessment on student learning: what can the students tell us? *Med Educ*. 2015;49(5):453-6.
27. 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):229-45.
28. 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):307-20.
29. 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):1492-7.
30. Dweck CS. Motivational Processes Affecting Learning. *Am Psychol*. 1986;41(10):1040-48.
31. Eva KW, Munoz J, Hanson MD, Walsh A, Wakefield J. Which factors, personal or external, most influence students' generation of learning goals? *Acad Med*. 2010;85(10 Suppl):S102-5.
32. Wood D, Bruner JS, Ross G. The role of tutoring in problem solving. *J Child Psychol Psychiatry*. 1976;17(2):89-100.
33. Ramani S, Post SE, Konings K, Mann K, Katz JT, Van der Vleuten C. "It's Just Not the Culture": A Qualitative Study Exploring Residents' Perceptions of the Impact of Institutional Culture on Feedback. *Teach Learn Med*. 2017;29(2):153-61.
34. Bloxham S, Campbell L. Generating dialogue in assessment feedback: exploring the use of interactive cover sheets. *Assess Eval High Educ*. 2010;35(3):291-300.



CHAPTER 3

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

Published as:

Schut, S., Heeneman, S., Bierer, B., Driessen, E., Van Tartwijk, J., & Van der Vleuten, C. (2020). Between trust and control: Teachers' assessment conceptualisations within programmatic assessment. *Medical Education*, 54 (6). 528-37. <https://doi.org/10.1111/medu.14075>

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.

INTRODUCTION

Interest in using assessment for learning is increasing in medical education and expectations of its benefits are high.¹ Programmatic assessment attempts to overcome the traditional dichotomy of assessment purposes as either formative or summative by proposing a continuum of assessment stakes.^{2, 3} 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.⁴ 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.^{5, 6} For this reason, using programmatic assessment to support learning remains challenging in practice.^{1, 7, 8}

Teachers appear to play a particularly powerful role in fulfilling or undermining the learning potential of programmatic assessment.⁷ 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.⁹ 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.⁴ The complex and overlapping interplay of assessment purposes, such as in low-stake assessments, adds to the already complicated assessment processes.^{10, 11} Consequently, programmatic assessment may challenge teachers' conceptualisations of assessment.

Following the description of Thomson¹² 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 Bain¹³ 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.¹³ 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.¹⁴ Furthermore, teachers' assessment conceptualisations are often informed by their personal assessment experiences rather than by educational theory or the institution's assessment policies.^{10, 12} These differences between beliefs and practices are especially likely to emerge when teachers encounter dual-purpose assessments,¹⁵ 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.^{1, 10, 16, 17}

The perspective of teachers within programmatic assessment is a missing component in the medical education literature.¹⁸ 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.

METHODS

We used a constructivist grounded theory approach^{19, 20} to explore teachers' assessment conceptualisations and assessment relationships within programmatic assessment.

Sample

An extreme case sampling strategy was employed to select unique research settings known to provide significant insights about programmatic assessment.²¹ 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.^{5, 22, 23}

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 (e.g., tutor, coach, physician advisor/mentor, lecturer, coordinator, preceptor/supervisor); (b) type of low-stake assessment (e.g., standardised in-course tests, essays, [research] assignments, direct observations), and (c) variable lengths of relationships with learners (ranging from brief encounters to longitudinal relationships).

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.

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 I 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.^{19, 24} 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 counterexamples. 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 25 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 and May 2019. Ethical approval was obtained from the Dutch Association for Medical Education Ethical Review Board (NVMOERB ref. 2018.7.4) and the Cleveland Clinic's Institutional Review Board (IRB ref. 18-1516).

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 counterexamples, and supported the process of code construction and data interpretation.

TABLE 2 Participant characteristics (n = 23)

CHARACTERISTIC	SETTING	
	A	B
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* in assessment system	3	
Physician advisor or mentor <i>Supports learners to interpret feedback, reflect on performance, monitor progress and develop portfolios</i>	4	5
Tutor or coach <i>Facilitator of PBL meetings, helps learners to construct subject matter-related objectives</i>	5	5
Course or module director <i>Responsible for design and delivery of courses or modules within curriculum</i>	3	10
Preceptor or supervisor <i>Responsible for guidance and supervision during clinical workplace-based assessment and learning</i>	2	5
Portfolio, progress or promotion committee member <i>Involved with portfolio reviews and shared responsibility for high-stake decision process</i>		3

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

* Most teachers performed multiple roles.

A: Physician-Clinical Investigator Programme, Faculty of Health, Medicine and Life Science, Maastricht University, the Netherlands

B: Physician-Investigator Programme, Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, Ohio, USA

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).

Conceptualisations of low-stake assessments

Stimulating and facilitating learning

Despite the differences in teachers' formal positions (e.g., 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)

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 influenced how teachers facilitated learning: teachers thought a more directive learning 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)

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)

Teachers' engagement with learners in assessment relationships

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.

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)

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)

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.^{2, 3} 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 Stiggins²⁵ 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 acquiring of 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-regulation²⁶⁻²⁸ 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.^{5-7, 29} 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.^{1, 30, 31} Moreover, learners themselves have voiced the importance of agency to enable the potential of using assessment for their learning.⁷ 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'³² 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.³³ 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.^{7, 33} 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.^{1, 29, 34-36} 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.³⁷ 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,³⁸ 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.⁴ 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,¹³ 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.¹³ 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.

Limitations

Our findings should be considered in the light of a number of limitations. First, this study included two unique implementations of programmatic assessment (i.e., 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.²⁰ Second, assessment is a complex interaction of learner, task, teacher and context characteristics,³⁹ which makes generalisations to other contexts challenging.⁴⁰ 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.

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.

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):309-17.
3. Schuwirth LW, Van der Vleuten CP. Programmatic assessment: From assessment of learning to assessment for learning. *Med Teach.* 2011;33(6):478-85.
4. Van der Vleuten CP, Schuwirth LW, Driessen EW, Dijkstra J, Tigelaar D, Baartman LK, et al. A model for programmatic assessment fit for purpose. *Med Teach.* 2012;34(3):205-14.
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):487-98.
6. Bok HG, Teunissen PW, Favier RP, Rietbroek NJ, Theyse LF, Brommer H, 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):654-63.
8. Harrison C, Wass V. The challenge of changing to an assessment for learning culture. *Med Educ.* 2016;50(7):704-6.
9. Uijtdehaage S, Schuwirth LWT. Assuring the quality of programmatic assessment: Moving beyond psychometrics. *Perspect Med Educ.* 2018;7(6):350-1.
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):442-67.
11. Black P, Wiliam 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 DA, editor. *Handbook of Research on Mathematics Teaching and Learning*. New York: Macmillan; 1992. p. 127-46.
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. *J Teach Educ.* 2011;27(1):210-20.
16. Rea-Dickins P. Understanding teachers as agents of assessment. *Lang Test.* 2004;21(3):249-58.
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):434-44.
19. Charmaz K. *Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis*. Thousand Oaks: Sage Publications, Inc.; 2006.
20. Corbin J, Strauss A. *Basics of Qualitative Research (3rd ed.): Techniques and Procedures for Developing Grounded Theory*. California: Thousand Oaks; 2008.
21. Creswell JW. *Educational Research: Planning, Conducting and Evaluating Quantitative and Qualitative Research*. Essex: Pearson Education Limited; 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):560-3.
 24. Watling C, Lingard L. Grounded theory in medical education research: AMEE Guide No. 70. *Med Teach.* 2012;34(10):850-61.
 25. Stiggins R. Two Disciplines of Educational Assessment. *Meas Eval Couns Dev.* 1993;26(1):93-104.
 26. Pratto F. On power and empowerment. *Br J Soc Psychol.* 2016;55(1):1-20.
 27. Heron J. Assessment revisited. In: Boud D, editor. *Developing Student Autonomy in Learning*. Second ed. London: Kogan Page Limited; 1981. p. 55-68.
 28. Schut S, Driessen E. Setting decision-making criteria: is medical education ready for shared decision making? *Med Educ.* 2019;53(4):324-6.
 29. 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):276-84.
 30. Harrison CJ, Konings KD, Schuwirth L, Wass V, Van der Vleuten CP. Barriers to the uptake and use of feedback in the context of summative assessment. *Adv Health Sci Educ Theory Pract.* 2015;20(1):229-45.
 31. Cilliers EJ, 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.
 32. 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):1092-9.
 33. 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):215-20.
 34. 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):1492-7.
 35. 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):307-20.
 36. Telio S, Regehr G, Ajjawi R. Feedback and the educational alliance: examining credibility judgements and their consequences. *Med Educ.* 2016;50(9):933-42.
 37. Meeuwissen SNE, Stalmeijer RE, Govaerts M. Multiple-role mentoring: mentors' conceptualisations, enactments and role conflicts. *Med Educ.* 2019;53(6):605-15.
 38. 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):162-5.
 39. Gipps CV. *Beyond testing: towards a theory of educational assessment*. London: The Falmer Press; 1994.
 40. Black P, Wiliam D. Lessons from around the world: how policies, politics and cultures constrain and afford assessment practices. *The Curriculum Journal.* 2005;16(2):249-61.

Appendix I: Interview guide

Topics and promoting questions

Main topics

- a. The concept of low-stakes assessment within a programmatic assessment system (if needed in comparison to high-stake assessment or previous assessment experiences)
- b. Roles and responsibilities (teacher/learner) in assessment.
- c. Interactions with learners in the context of programmatic assessment.
- d. Relationship between assessment and learning.

HOW – WHY – EXAMPLES

Topic questions

- What do low-stake assessments mean to you? Could you give a description these assessments?
- How do you use low-stake assessment in practice? Could you elaborate on that assessment practice? Is this different from high-stake assessments? Could you elaborate on that?
- How do you value low-stake assessment?
- Could you describe the consequences based on learners' performance on the assessment task? How did you feel about these consequences?
- Could you describe your role and responsibility within the assessment program?
- What do you believe is your role and responsibility with low-stake assessments? How do you feel about that? Could you give an example?
- What do you believe is the learner's role and responsibility with low-stake assessments? How do you feel about that?
- Did these roles and responsibilities changed when programmatic assessment was implemented?
- Could you describe your relationship to the learner in programmatic assessment? How do you interact with learners concerning assessment and learning? How do you feel about that relationship? Do you experience any differences?
- How do you value assessment and learning? What do you consider the relationship between assessment and learning? How do you feel about that? Did that change?

General prompting questions

- Can you give an example? Can you give a counter example?
- Can you elaborate more on this?
- Why is that?
- Can you explain this?
- How do you feel about this?

Room for not-discussed themes



CHAPTER 4

Understanding the influence of teacher–learner relationships on learners’ assessment perception

Published as:

Schut, S., Van Tartwijk, J., Driessen, E., Van der Vleuten, C., & Heeneman, S. (2020).
Understanding the influence of teacher-learner relationships on learners' assessment
perception. *Advances in Health Sciences Education*, 25 (2). 441-56.
<https://doi.org/10.1007/s10459-019-09935-z>

ABSTRACT

PURPOSE Low-stakes assessments are theorised to stimulate and support self-regulated learning. They are feedback-, not decision-oriented, and should hold little consequences to a learner based on their performance. The use of low-stakes assessment as a learning opportunity requires an environment in which continuous improvement is encouraged. This may be hindered by learners' perceptions of assessment as high stakes. Teachers play a key role in learners' assessment perceptions. By investigating assessment perceptions through an interpersonal theory-based perspective of teacher–learner relationships, we aim to better understand the mechanisms explaining the relationship between assessment and learning within medical education.

METHODS First, twenty-six purposefully selected learners, ranging from undergraduates to postgraduates in five different settings of programmatic assessment, were interviewed about their assessment task perception. Next, we conducted a focussed analysis using sensitising concepts from interpersonal theory to elucidate the influence of the teacher–learner relationship on learners' assessment perceptions.

RESULTS The study showed a strong relation between learners' perceptions of the teacher–learner relationship and their assessment task perception. Two important sources for the perception of teachers' agency emerged from the data: positional agency and expert agency. Together with teacher's communion level, both types of teachers' agency are important for understanding learners' assessment perceptions. High levels of teacher communion had a positive impact on the perception of assessment for learning, in particular in relations in which teachers' agency was less dominantly exercised. When teachers exercised these sources of agency dominantly, learners felt inferior to their teachers, which could hinder the learning opportunity.

CONCLUSIONS To utilise the learning potential of low-stakes assessment, teachers are required to stimulate learner agency in safe and trusting assessment relationships, while carefully considering the influence of their own agency on learners' assessment perceptions. Interpersonal theory offers a useful lens for understanding assessment relationships. The Interpersonal Circumplex provides opportunities for faculty development that help teachers develop positive and productive relationships with learners in which the potential of low-stakes assessments for self-regulated learning is realised.

INTRODUCTION

Creating a learning opportunity with the use of assessment, which can benefit self-regulated learning,¹ has gained greater prominence in medical education literature.² It remains a major challenge because assessments designed to inform and support learners' progress are often perceived to be high-stakes, summative assessments by learners.³⁻⁵ Utilising the learning potential of assessment requires a better understanding of the mechanisms that explain the relationship between assessment and learning. The current international study involved a focused analysis of data collected as part of a larger research project on the learning potential of programmatic assessment, specifically low-stakes assessments.⁶ One of the findings reported in this previous paper was that the teacher-learner relationship was an important factor influencing learners' perception of assessment stakes. In the study reported in this paper, we explored how teacher-learner relationships influence this perception and how the influence of these relationships is negotiated in the context of assessment.

Low-stakes assessments are designed to support learning. All kinds of assessment instruments can be used as low-stakes assessments, ranging from multiple choice tests to direct observations of performance in the clinical workplace. They differ from high-stakes assessments in their consequences; low-stakes assessments are feedback-oriented and should hold little consequences to a learner based on their performance. In programmatic assessment they have a dual purpose: as individual assessment to support learning and to provide feedback, and information for learner and teacher, but when aggregated they are used to inform make high-stakes decisions.

Learners' perceptions of assessment influence how they learn,⁷ and teachers play a powerful role in those perceptions.^{2, 6} Teaching happens through human interaction, and therefore the characteristics of teachers' interaction and relationships with learners can make a substantial difference to the kind of learning environment they create.⁸⁻¹⁰ Research outside the context of medical education has shown that learners' perceptions of teacher-learner relationships have a large impact on the quality of learners' motivation, as well as on how they engage in various tasks.¹¹⁻¹³ Furthermore, learner support has been identified and highlighted as key when we aim to use assessment as a learning opportunity.^{14, 15} However, there is still a significant gap in the medical education literature with respect to the effects of teacher-learner relationships on the use and uptake of assessment information,^{2, 8, 10} and therefore the potential of using assessment to support learning. Our aim with this study is to develop a better understanding of how teacher-learner relationships influence the learning potential of low-stakes assessments.

To understand the influence of teacher-learner interactions and relationships on learners' assessment perception, interpersonal theory is pivotal. The theory states that to describe interpersonal communication, two independent dimensions are both sufficient and necessary: agency and communion.¹⁶⁻¹⁹ Applied to education,²⁰ Teacher Agency describes the level of teacher influence in the teacher-learner interaction and relationship. Teacher Communion refers to the level of warmth or friendliness a teacher communicates in interactions.^{18, 20-22}

CHAPTER 4

Interpersonal circumplex models combine these two independent dimensions in one framework, indicating that the interpersonal meaning of behaviour in interaction and positions in relationships can be described as a combination of both dimensions (Fig. 1). The principle of complementarity¹⁹ suggests that the interpersonal positions of two people in interactions or relationships are not random but occupy a similar position on the dimension of communion (e.g., high teacher communion invites high learner communion and vice versa) and an opposite position of the dimension of agency (e.g., high teacher agency will invite low learner agency and vice versa).

In the literature, two levels of time are distinguished when describing interpersonal perceptions.²³ Interpersonal perceptions of behaviour and interactions are positioned at the moment-to-moment: the micro-level. Relationships can only be described at a longer period of time: the macro-level. For instance, when a preceptor demands a resident to pay attention during rounds, it is likely that this behaviour is perceived as dominant (micro-level). A physician who has full attention during hand-over will be perceived as having a high agency position in relationship to colleagues (macrolevel). Micro and macro-level perceptions mutually influence each other. For instance, supervisors with a high agency position in relationships with residents, will need to ask for attention less often.

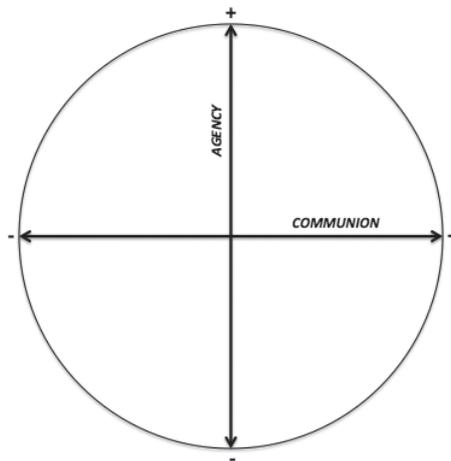


FIGURE 1. The Interpersonal Circumplex (e.g., Gurtman¹⁸, Wiggins²⁷).

Interpersonal theory has been used extensively to investigate teacher–learner interactions and relationships in secondary education,¹³ and in the supervision process within research projects in higher education and PhD projects.^{24, 25} Studies have generally shown positive effects of combinations of teachers’ agency and communion on cognitive and affective learning outcomes.^{11–13} Teacher support and caring (communion) has been proven to be critically important

for learners' interest in academics,²⁶ their efficacy and engagement in self-regulated learning, and the decrease of their anxiety about task engagement.²⁷ Studies have shown positive influences of teacher communion on feedback receptivity in the context of master thesis supervision²⁴ and PhD supervision.²⁸ Furthermore, teacher support influences learners' achievement goals, specifically by promoting the adoption of mastery goals (defined as goals in which the emphasis is on learning, effort, and improvement).²⁹ This is likely to stimulate long-term and high-quality involvement in learning and is associated with a wide range of motivation-related variables that are necessary mediators for self-regulated learning.⁷ Unsurprisingly, learners care about their relationships with teachers and respond with greater engagement and effort when they believe teachers care and are supportive.^{30, 31} Far less attention has been paid to the relation of teachers' agency with learners' achievement goals³² and assessment task perception, which is of interest in the current study, especially given the hierarchical nature and location of power within any assessment relationship.^{33, 34} The context of medical education further complicates the struggle of teachers with how much autonomy to grant learners, given the responsibility and need for high-quality patient care within the medical workplace.^{14, 35, 36}

In the present study, the agency and communion dimensions are used to describe and analyse learners' reflections on interpersonal relationships with their teachers in the assessment context. The following research question guided the study: How are learners' perceptions of the interpersonal relationship with teachers in the assessment environment related to learners' assessment perceptions from early undergraduate medical education to postgraduate training?

METHODS

We used an open and qualitative approach to data gathering and analysis, inspired by constructivist grounded theory^{37, 38} to explore learners' experiences with low-stakes assessments in the context of programmatic assessment. Learners' perceptions of the relationship with their teachers, particularly their sense of agency in this relationship, was identified as a key variable for understanding learners' assessment perception.⁶ To further delve into this relationship aspect of assessment perception, we turned to interpersonal theory¹⁹ where two dimensions used to conceptualise interpersonal perceptions (agency and communion). These dimensions were used as existing sensitising concepts to help explain the underlying mechanism and enhance our understanding of the influence of interpersonal relationships on the perception and use of low-stakes assessments. We thus used a focussed analysis to examine all data relating to participants' comments that pertained to their teachers' interpersonal behaviour or to their interactions or relationships with their teachers, using an interpersonal theory lens to enhance the explanatory power of the emerging theory.³⁷ For the purpose of the current study, teachers are defined as 'all significant others' whom learners encounter within their assessment experiences. The assessment practices are described in detail elsewhere.⁶

Data collection

This study was conducted in three institutes: Cleveland Clinical Lerner College of Medicine (USA), Maastricht University (Netherlands), and Dalhousie University (Canada). The context of medical education varies over time and learners encounter teachers with different educational roles in various assessment environments. Learners move from a highly structured academic learning context in the pre-clinical years, to the more open, and unstructured clinical context of undergraduate and postgraduate medical training. The latter demands a more active, self-steering attitude from learners over time³⁹, which potentially sets different requirements for the relationship with their teachers. We therefore purposefully selected learners experiencing different assessment practices from preclinical and clinical phases: preclinical undergraduate education (setting A, Netherlands and USA); clinical undergraduate education, also known as clerkships (setting B, Netherlands); and post-graduate medical education, also known as residency programs (setting C, Canada, Netherlands). This resulted in a sample from five different assessment programmes from three different institutes. E-mails inviting learners to participate in one-to-one interviews were sent by local faculty members. A convenience sampling approach was taken based on learners' availability at pre-determined times. A total of 26 learners participated in individual, semi-structured interviews. An overview of participants' characteristics can be found in Table 1.

TABLE 1. Participants' characteristics

Characteristics		N=26
Programme		
A1	Year 1 & 2 of the 5-year graduate entry programme Physician-Clinical Investigator at Cleveland Clinical Lerner College of Medicine of CWRU, Ohio, USA	5
A2	Year 2 of the 4-year graduate-entry programme Physician-Clinical Investigator at the Faculty of Health, Medicine and Life Science, Maastricht University, the Netherlands	6
B1	The 12 weeks clinical rotation Family Medicine during the undergraduate programme Medicine at Faculty of Health, Medicine and Life Science, Maastricht University, the Netherlands	4
C1	Year 2 of the 2-year Family Medicine residency programme at Dalhousie University Department of Family Medicine, Canada	6
C2	End year 1 and 3 of the 3-year family Medicine residency programme at Maastricht University Medical Centre, Maastricht	5
Learning and assessment setting		
	Preclinical training	11
	Clinical training	15
Gender		
	Female	18
	Male	8

Open-ended questions were posed by one interviewer (SS), who asked participants to describe (1) their general perceptions of programmatic assessment, (2) their different assessment experiences, (3) how they perceived the assessment stakes and what influenced these stakes. When participants mentioned the role and influence of the teacher, follow-up questions were used to prompt a more detailed reflection on how teachers influenced assessment stakes and how learners perceived the relationship. See Appendix I for the Topic Guide for Interviews. Data were collected between April 2016 and November 2016. Participants received a small compensation (a \$/€10 gift card). Ethical approval was obtained from the Dutch Association for Medical Education Ethical Review Board (NVMO-ERB668 on 01/03/2016), the Dalhousie Health Sciences Research Ethics Board (REB#2016-3882 on 25/07/2016) and the Cleveland Clinic Institutional Review Board (IRB#16-1261 on 21/09/2016). The application and approval allowed for the iterative process of data collection and analysis following the open, and explorative nature of our research project. All participants signed the consent form in which they were informed about the nature of our project and analyses, before and after the interview.

Data analysis

Verbatim transcripts of the interviews were made and analysed using a constant comparison approach, which comprises simultaneous coding and analysis of data in order to develop and refine concepts and explore their inter-relationship.³⁷ The focussed analysis combined coding that was guided a priori by awareness of interpersonal theory with inductive coding that emerged from the data. Firstly, we identified the points in the transcripts where participants mentioned teachers in their assessment and learning environment. All fragments were re-read to identify whether or not a learner reflected on the behaviour of the teacher, past interpersonal experiences, or the nature of the relationship. These points were coded as ‘Interpersonal Perceptions (IP) fragments’: one or a sequence of sentences relating to the behaviour of the teacher, past interpersonal experiences or the nature of the relationship as perceived by the learner. Secondly, we identified the points in the transcripts where participants mentioned reasons, motivations, goals, and ambitions to engage with their assessment task and how they perceived the assessment task in stakes and learning value. These points were distinguished within the transcripts as learners’ Assessment Task Perception (ATP) fragments. All fragments were re-examined and coded according to the perceived stakes (low- to high-stakes) and the perceived value the assessment has to guide (further) learning. The initial template was used to re-examine all identified fragments and their relationship within each transcript by the primary investigator (SS), allowing additional codes to modify and refine the template. Memos and comments were added to further specify and elaborate code meaning. This helped us understand the richness of the data and the mechanisms involved in interpersonal perceptions, as well as the influence the perceived teacher-learner relationship had on learners’ assessment task perception. A second investigator (SH) independently analysed and re-coded the fragments of the first five interviews. Both investigators then met to compare their interpretations and resolve disagreements through discussion. To further enhance the rigour of the analysis, examples and counter examples were reviewed by a third investigator, who is considered an expert in teacher-learner relationships

and teacher education in general (JvT). The final template was discussed with the whole research team and all fragments were reread (SS) to ensure no relevant information was missed.

We acknowledge that the data as well as the interpretations and meaning we ascribed to these data are co-constructed by interactions with the participants. To prevent biases as much as possible, we collected data from multiple sources and brought together a multidisciplinary research team for analysis: SS and ED have a background in educational sciences; CvdV in psychology, JvT in sociology of education, and SH in biomedical sciences. JvT is not directly involved in medical education. Furthermore, to avoid confirmation bias in our interpretations, ED, SH and CvdV brought an outsider perspective to interpersonal theory.

RESULTS

Our analysis of the data confirmed that the agency and communion dimensions provide an adequate frame of reference for enhancing our understanding of learners' perceptions of assessment stakes. The results of these analyses will be described underneath, structured according to the two dimensions, and supported with illustrative examples.

Accepting teachers' agency as legitimate

Positional agency

The first source of teachers' agency came with their authority position which is referred to as "positional agency": agency that stems from holding social positions such as rank or an authority role. Most learners indicated that in the assessment process, they often felt subordinate to teacher's authority position: they emphasised that teachers had the power to determine content, performance criteria, and consequences, as well as to assess and qualify learners' performance, based on their formal position as teacher/assessor. Learners especially felt they depended on the teacher as an authority when assessment criteria were perceived as subjective judgments (as contrasted to standardised tests), for example in the case of performance judgements based on work-based observations. Although the assessment context was low-stakes, learners were aware that all encounters could influence or might

determine the high-stakes evaluation. Even when the teacher was perceived as supportive, there was still a feeling that "at the end of the day" (C2.1), it was the decision of the teacher that mattered. Furthermore, there was a general feeling that assessments and evaluations from staff members were valued more highly when evaluating learners' progress, than for example self- or peer assessments.

It's known that the assessments from the faculty do carry more weight. Because they come from faculty and their position, I think part of that is just that authority position, that their assessment does carry more weight. (A1.12)

Not all positional agency was necessarily accepted as legitimate. In several situations, positional agency led learners to comply with the demands or requirements of the teacher, even when they did not believe these had any value or meaning for them. This tended to alienate them from their learning process and made the whole assessment experience less meaningful for learning.

So that makes me feel overpowered. I think, well, I could explain my point of view, but when he believes it is like that... well I can't change that [...], in the end I'm still dependent on what he believes. (B1.15)

Some learners, especially within undergraduate medical education, were particularly focused on teachers' positional agency. These learners looked to their teacher for direction, valued teachers' judgement highly, often wanted to impress teachers, and consequently raised the stakes themselves. They looked for approval and were willing to comply with the normative beliefs of teachers.

The higher up in hierarchy, the higher the stakes become and the more value I attribute to it. If a resident says I'm not doing okay, I think, well, you are just a resident yourself huh, but when a professor says something like that, then I think, oh no. That obviously matters more. (B1.18)

Furthermore, some learners expressed a need for an 'objective' qualification of their performance from a formal authority figure for summative, accountability purposes. Although this led to a high-stakes assessment experience, learners could still value the learning opportunity.

Expert agency

The second source of teachers' agency that enhanced teachers' influence on learners' assessment perception was expertise. When teachers were perceived as knowledgeable, learners were more willing to accept this base of influence as legitimate and to use assessment as a learning opportunity: I have grown respect for my particular preceptor, realised over time just how good he is at what he does and so now when he gives me advice, I'm very focused on what that is and how to apply it. (A1.11). The value of teachers' expertise did not depend solely on medical expertise, but also on knowledge and experience as an educator. Although teachers' medical expertise related to the perceived relevance of assessment and learners' feedback receptivity, trust and safety also related to how knowledgeable the teacher was of the objectives and requirements of medical training and the assessment system.

I think very highly of my preceptor, but as a surgeon that is. If I needed surgery, he would be my go-to. But as a teacher not so much. I don't care so much what he thinks, he hardly knows what I'm supposed to do or learn, and he has no idea about how I function in practice. He just doesn't have that expertise. (C2.3)

Communion: engaging in safe and personal relationships

Recognition and engagement were important for a personal relationship to become established, which contributed to a safe learning environment and the opportunity to use assessment as a learning experience (high communion). Learners seemed more willing to show their weaknesses when they felt that teachers took a genuine interest in their learning process and invested time in understanding and getting to know them. Learners emphasised that sharing personal life experiences with the teacher made them feel the teacher was approachable, which created an assessment environment in which they allowed themselves to be open and to learn from low-stakes assessments.

I think also because you spend so much time together, you see each other the entire day, you go on house visits together, and I, we had lunch with his family at home every day. You just really, you get to know each other. It makes it all feel less high stake. (B1.15)

In contrast, when learners perceived teachers to be low on communion, they seemed more intent on giving a positive image of themselves and avoiding losing face. When learners encountered teachers who didn't invest time, didn't try to get to know them, or showed little interest in their development, assessments rarely led to a learning opportunity.

I think it's more high-stake. The interactions with these staff people, you don't spend enough time with them, they don't care, you're just another learner [...], it makes you more nervous, it's more stressful. (C1.8)

Mostly with short-term encounters, and often with preceptors or supervisors from a different rotation or clerkship, learners felt like a burden on the available time of the teacher. They felt less inclined to ask for feedback and generally experienced the assessment as high stakes. Learners indicated that time spent together was an important requirement for an honest and more holistic judgment by the teacher, which they were then more willing to accept.

Yeah, they know my style, they also know my weaknesses and strengths, I think because it's such a long-term relationship, I feel like it's easier for me to come to them and say I'm struggling with this. [...] It helps to become more vulnerable, and I get to be really honest and look for their guidance. Whereas I might not really do that in any of the other settings because I don't want to be judged by the preceptors who don't know me really well, based on one single interaction. (C1.8)

Combining the dimensions: effective assessment relationships to facilitate learning

When learners perceived teachers to be high on agency but not showing dominance at the behaviour level, and to be combining it with friendly behaviour (high communion), it was more

likely that low-stakes assessment was used as a learning opportunity. These teachers (high agency–high communion) were described as showing a genuine interest in learners’ objectives, strengths and weaknesses. They also allowed learners to influence or determine assessment objectives. In relationships that were characterised by such teacher behaviour, learners felt comfortable and felt more agency in the assessment process. These relationships enabled assessment to be used as a learning opportunity, which is an illustration of the principle of complementarity.

I think it was the best because, she, I felt more that we were like on equal parts [...]. So it was that combined discussion and continued learning that was going on. It wasn't like I am up here on the power differential and talking down to you and this is what you have done, it was working together. (C1.9)

Teachers with high positional agency whose behaviour was not only perceived as dominant but also as uninterested or cold (low communion) caused learners to feel more dependent. In general, these teachers (high agency–low communion) were perceived as less approachable and learners had no influence on the assessment or feedback, making all assessments high-stakes. This translated in the need for teachers’ approval and in less autonomy over their own learning process: “*I will still ask some questions, but it's more like, gee, just tell me what you want from me.*” (A2.19)

DISCUSSION

In this paper, interpersonal theory provided a useful lens to examine and understand the associations between the characteristics of teacher–learner relationships and learners’ assessment perceptions. Two important sources for teachers’ agency emerged from the data: positional agency and expert agency. When teachers make these sources of agency prominent by behaving dominantly, learners felt subordinate to their teachers, which hindered assessment being used as a learning opportunity. High levels of teacher communion, in particular in relations in which teachers’ agency was exercised less dominant, had a positive impact on the perception of assessment for learning. What this paper adds is a unique understanding of how these relationships influence learners’ assessment experiences, what sources for agency are involved in assessment relationships and how agency is exercised and negotiated in such a context.

The ambitions for self-regulated learning challenge the traditional assessment model in which teachers exercise unilateral intellectual authority over the assessment process.³³ What counts as legitimate knowledge is in practice often still framed by the teacher and the summative assessment demands.⁴⁰ This automatically situates the teacher high on agency, which was also found in our study. Competency-based medical education and the use of portfolios, potentially creates even more dependency on teacher’s agency, because such assessments require teachers’ expert judgment.¹ Based on our findings, we argue that in order to utilise the learning potential of assessment, the assessment process requires more learner’s agency and a shift to assessment as a process of co-inquiry to determine what counts as legitimate. Teachers with positional and/or expert agency who do not show dominance, and who engage in personal, trusting relationships,

are most likely to create safe learning environments in which learners feel they have such agency. Such teachers are able to provide rules and structure when needed, for example when patient safety is at stake, while avoiding restrictions on learner autonomy. As learners proceed through different phases of medical training, the need for rules and structure change, and teachers need to be able to adapt to those needs when striving to benefit self-regulated learning, by entrusting and empowering learners gradually.

The communion dimension and the complementary principle that elicits sameness (e.g., high teacher communion invites high learner communion and vice versa) explained the conditions for a safe and personal relationship to develop, in which assessment could be experienced and used as a learning opportunity. Time was considered an important prerequisite for high-communion relationships to develop, which is also highlighted by other scholars.^{2, 41, 42} Our results showed that during medical training, this is not always possible. Caution should be taken with the influence of short-term interpersonal relationships on learners' assessment perception. Medical educators should support learners' resilience for dealing with short-term relationships and the impact these have on assessment and learning and create awareness amongst teachers that assessment should be approached with more caution when a relationship is lacking.

Practical implications of this study appertain to teacher professionalisation. We argue it is worthwhile to consider how learners perceive teachers interpersonally when one is aiming to use assessment to create a learning opportunity. The 'Interpersonal Circumplex'¹⁸ was found useful to understand the influence of teacher-learner relationships on learners' assessment perception, and which type of interactions led learners to use the assessment experience as a learning opportunity. Given that interactions are seen as the building blocks of relationships,^{13, 17, 21} knowledge of those interactions and the influences these have on the use and uptake of assessment, are useful when implementing the use of assessment for learning. This is particularly the case for the diagnosis of (problematic) interactions or undesirable assessment perceptions, but also for the design of effective interventions that stimulate a more effective relationship between teachers and learning in the context of assessment. We argue that the Interpersonal Circumplex¹⁸ could be used to create awareness amongst teachers on how they influence learners' assessment experiences and how they can alter these interactions and engage in effective relationships, as has been done in the domain of teacher education.⁴³ The situational context of medical education and assessment is important here. Medicine's learning culture might constrain its own teachers for building relationships that make meaningful contributions to learners' development, due to the important values of independence and autonomy that characterise medicine's culture.^{9, 14, 44, 45} Furthermore, we acknowledge that teachers might face significant dilemmas and struggle with the tension that exists between their supportive role in monitoring and facilitating learners' development and their judgmental responsibility as assessor of learners' performance and achievement,⁴⁶⁻⁴⁸ especially given the requirements of their role within medical education to warrant the quality of future doctors, who will be required to provide safe and effective health care.¹⁴

Limitations

This study has several limitations. Firstly, the approach adopted here is that all analyses are based on learners' perceptions only. Although many studies have advocated that how learners perceive teachers' behaviour may be more informative than more objective measures,⁴⁹ this could have influenced our results. High correspondence between different respondents' perceptions of the same type of interactions leading to the same assessment perceptions did strengthen our interpretations and conclusions of the value of the interpersonal dimensions to explain, understand and eventually steer assessment perceptions and the utilising of the potential of programmatic assessment. We do stress the importance of including teachers' perceptions and observations of interpersonal communication to further build this theory. Secondly, learners participated voluntarily which potentially creates a selection bias. It might very well be that these learners' reflections might not represent those of other learners. Finally, the data were collected within different institutes and phases of medical education, limiting the number of interviews per setting. Especially given the cultural nature of concepts like interpersonal relationships, assessment, and self-regulated learning, we stress the importance of further testing the value of interpersonal theory within different disciplines of medical education in various countries. Furthermore, a study using a longitudinal design could shed more light on the development of learners' perception of teacher-learner relationships in the context of assessment.

CONCLUSION

The 'Interpersonal Circumplex'¹⁸ was found useful to understand the influence of teacher-learner relationships on learners' assessment perception, and which type of interactions led learners to use the assessment experience as a learning opportunity. We argue that although teacher-learner relationships in the context of assessment are essentially unequal, their purpose should be to reduce this inequality when one is aiming to create an assessment environment that facilitates self-regulated learning. Ideally, the teacher would help learners' competency development, and their ability to set and achieve useful goals, without trying to oppose or override learners' agency. Medical education might still be far from establishing a true learning culture,^{2,15} but the Interpersonal Circumplex model could be instrumental to help us to move towards a culture that emphasises learning and that supports continuous improvement. Faculty development focussing on interpersonal relationships should be given priority, given the challenges teachers face in overcoming traditional power relationships in assessment and their responsibly to patient safety within the context of medicine specifically. However, teachers have the power to enable the use of assessment as a learning opportunity and should receive support in fulfilling that promise.

REFERENCES

1. Van der Vleuten CP, Schuwirth LW, Driessen EW, Dijkstra J, Tigelaar D, Baartman LK, et al. A model for programmatic assessment fit for purpose. *Med Teach*. 2012;34(3):205-14.
2. Watling CJ, Ginsburg S. Assessment, feedback and the alchemy of learning. *Med Educ*. 2019;53(1):76-85.
3. Bok HG, Teunissen PW, Favier RP, Rietbroek NJ, Theyse LF, Brommer H, et al. Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Med Educ*. 2013;13:123.
4. 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):487-98.
5. 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):276-84.
6. 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):654-63.
7. Ames C. Classrooms: Goals, structures, and student motivation. *J Educ Psychol*. 1992;84(3):261-71.
8. Haidet P, Stein HF. The Role of the Student-Teacher Relationship in the Formation of Physicians : The Hidden Curriculum as Process. *J Gen Intern Med*. 2006;21(S1):S16-S20.
9. Ramani S, Post SE, Konings K, Mann K, Katz JT, Van der Vleuten C. "It's Just Not the Culture": A Qualitative Study Exploring Residents' Perceptions of the Impact of Institutional Culture on Feedback. *Teach Learn Med*. 2017;29(2):153-61.
10. Telio S, Ajjawi R, Regehr G. The "educational alliance" as a framework for reconceptualizing feedback in medical education. *Acad Med*. 2015;90(5):609-14.
11. Den Brok P, Brekelmans M, Wubbels T. Interpersonal Teacher Behaviour and Student Outcomes. *Sch Eff Sch Improv*. 2004;15(3-4):407-42.
12. Wentzel KR. Motivating Students to Behave in Socially Competent Ways. *Theory Pract*. 2003;42(4):319-26.
13. Wubbels T, Brekelmans M. Two decades of research on teacher–student relationships in class. *Int J Educ Res*. 2005;43(1-2):6-24.
14. Eva KW, Bordage G, Campbell C, Galbraith R, Ginsburg S, Holmboe E, et al. Towards a program of assessment for health professionals: from training into practice. *Adv Health Sci Educ Theory Pract*. 2016;21(4):897-913.
15. 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):1492-7.
16. Bakan D. *The duality of Human Existence: Isolation and Communication in Western Man*. Boston, MA: Beacon Press; 1966.
17. Fournier MA, David DSM, Zuroff DC. Origins and Applications of the Interpersonal Circumplex. In: Horowitz LM, Strack S, editors. *Handbook of Interpersonal Psychology: Theory, Research, Assessment, and Therapeutic Interventions*. Hoboken, New Jersey: John Wiley & Sons, Inc.; 2011. p. 57-73.
18. Gurtman MB. Exploring Personality with the Interpersonal Circumplex. *Soc Person Psychol Compass*. 2009;3(4):601-19.
19. Horowitz LM, Strack S, editors. *Handbook of Interpersonal Psychology: Theory, Research, Assessment, and Therapeutic Interventions*. Hoboken, New Jersey: John Wiley & Sons, Inc.; 2011.
20. Wubbels T, Brekelmans M, Mainhard T, Den Brok P, Van Tartwijk J. Teacher–student relationships and

- student achievement. . In: Wentzel KR & Ramani GB, editors. *Handbook of Social Influences in School Contexts: Social-Emotional, Motivation, and Cognitive Outcomes* New York: Routledge; 2016. p. 127-42
21. Pennings HJM, Brekelmans M, Claessens LCA, Van Tartwijk J, Sadler P, Van der Want AC. Interpersonal adaptation in teacher-student interaction. *Learn Instr.* 2018;55:41-57.
 22. Wiggins JS. *The five-factor model of personality: theoretical perspectives.* New York: Guilford Press; 1996.
 23. Wubbels T, Den Brok P, Van Tartwijk J, Levy J. Introduction to interpersonal relationships in education. In: Wubbels T, Den Brok P, Van Tartwijk J, Levy J, editors. *Interpersonal relationships in education. An overview of contemporary research.* Boston: Sense Publishers; 2012. p.1-15.
 24. De Kleijn RAM, Meijer PC, Pilot A, Brekelmans M. The relation between feedback perceptions and the supervisor–student relationship in master's thesis projects. *Teach High Educ.* 2013;19(4):336-49.
 25. Mainhard T, Van der Rijst R, Van Tartwijk J, Wubbels T. A model for the supervisor–doctoral student relationship. *High Educ.* 2009;58(3):359-73.
 26. Wentzel KR. Relations of Social Goal Pursuit to Social Acceptance, Classroom Behavior, and Perceived Social Support. *J Educ Psychol.* 1994;86(2):173-82.
 27. Ryan AM, Patrick H. The Classroom Social Environment and Changes in Adolescents' Motivation and Engagement during Middle School. *Am Educ Res J.* 2001;38(2):437-60.
 28. Ives G, Rowley G. Supervisor selection or allocation and continuity of supervision: PhD students' progress and outcomes. *High Educ Stud.* 2005;30(5):535-55.
 29. Turner JC, Gray DL, Anderman LH, Dawson HS, Anderman EM. Getting to know my teacher: Does the relation between perceived mastery goal structures and perceived teacher support change across the school year? *Contemp Educ Psychol.* 2013;38(4):316-27.
 30. Urdan T, Schoenfelder E. Classroom effects on student motivation: Goal structures, social relationships, and competence beliefs. *J Sch Psychol.* 2006;44(5):331-49.
 31. Wentzel KR. Social-Motivational Processes and Interpersonal Relationships: Implications for Understanding Motivation at School. *J Educ Psychol.* 1999;91(1):76.
 32. Mainhard T. Liking a tough teacher: Interpersonal characteristics of teaching and students' achievement goals. *Sch Psychol Int.* 2015;36(6).
 33. Heron J. Assessment revisited. In: Boud D, editor. *Developing Student Autonomy in Learning.* Second ed. London: Kogan Page Limited; 1981. p. 55-68.
 34. Reynolds M, Trehan K. Assessment: A critical perspective. *High Educ Stud.* 2000;25(3):267-78.
 35. Hauer KE, Ten Cate O, Boscardin C, Irby DM, Iobst W, O'Sullivan PS. Understanding trust as an essential element of trainee supervision and learning in the workplace. *Adv Health Sci Educ Theory Pract.* 2014;19(3):435-56.
 36. Watling CJ, LaDonna KA. Where philosophy meets culture: exploring how coaches conceptualise their roles. *Med Educ.* 2019;53(5):467-76.
 37. Corbin J, Strauss A. *Basics of Qualitative Research (3rd ed.): Techniques and Procedures for Developing Grounded Theory.* California: Thousand Oaks; 2008.
 38. Watling CJ, Lingard L. Grounded theory in medical education research: AMEE Guide No. 70. *Med Teach.* 2012;34(10):850-61.
 39. Deketelaere A, Kelchtermans G, Struyf E, De Leyn P. Disentangling clinical learning experiences: an exploratory study on the dynamic tensions in internship. *Med Educ.* 2006;40(9):908-15.
 40. Pryor J, Crossouard B. A socio-cultural theorisation of formative assessment. *Oxf Rev Educ.* 2008;34(1):1-20.

CHAPTER 4

41. Ramani S, Konings KD, Mann KV, Pisarski EE, Van der Vleuten CPM. About Politeness, Face, and Feedback: Exploring Resident and Faculty Perceptions of How Institutional Feedback Culture Influences Feedback Practices. *Acad Med.* 2018;93(9):1348-58.
42. Sargeant J, Lockyer J, Mann K, Holmboe E, Silver I, Armson H, et al. Facilitated Reflective Performance Feedback: Developing an Evidence- and Theory-Based Model That Builds Relationship, Explores Reactions and Content, and Coaches for Performance Change (R2C2). *Acad Med.* 2015;90(12):1698-706.
43. Van Tartwijk J, Veldman I, Verloop N. Classroom management in a Dutch teacher education program: a realistic approach. *J Teach Educ.* 2011;22(2):169-84.
44. Harrison C, Wass V. The challenge of changing to an assessment for learning culture. *Med Educ.* 2016;50(7):704-6.
45. Shepard LA. The Role of Assessment in a Learning Culture. *Educ Res.* 2016;29(7):4-14.
46. Rea-Dickins P. Understanding teachers as agents of assessment. *Lang Test.* 2004;21(3):249-58.
47. Watling C. The uneasy alliance of assessment and feedback. *Perspect Med Educ.* 2016;5(5):262-4.
48. Wiliam D. Embedded formative assessment. Bloomington, IN: Solution Tree Press; 2011.
49. Shuell TJ. Toward an Integrated Theory of Teaching and Learning. *Educ Psychol.* 2010;28(4):291-311.

Appendix I: Interview guide

Topics and promoting questions

1. Low-stake assessment

- a. Different low-stake assessment experience (if needed in comparison to high-stake assessment)
- b. Information and feedback fostering all learning from assessment
- c. Role of the teacher, assessor, supervisor

HOW—WHY—EXAMPLES

General prompting questions

- Can you give an example? Can you give a counter example?
- Can you elaborate more on this?
- Why is that?
- Can you explain this?
- How do you feel about this?

Topic A

- Could you describe the stakes you perceive in that assessment? Do you have other experiences?
- What influenced the stakes involved?
- Does this apply for all assessment experiences (good, bad, low-stakes, high-stakes, positive, negative, etc.)
- Could you describe the consequences related to your performance on the assessment task?
- Can you tell us about the emotions related to these consequences you experienced?
- How did you feel about these consequences?
- What or who decided on these consequences?

Topic B

- Could you tell us something about the information you gain from assessment tasks that benefit your learning? Do you have counter examples?
- Could you describe assessment tasks or learning activities that provided you with information concerning your progress or development? Do you have counter examples?
- Why/how did you use this information? Do you have counter examples?
- Did you have control about assessment, feedback, time, content, etc.?

CHAPTER 4

Topic C

- How would you describe this relationship with your assessor/teacher/supervisor?
- How do you experience this relationship?
- Can you describe the role of your teacher, assessor, supervisor, etc. more elaborately? Do you experience any differences?
- How long did this relationship last? Did that influence the assessment experience or your relationship? Could you describe how? Do you have counter examples?
- What does this mean for the assessment experience and stakes involved?
- Are there other examples of relationships that influenced your assessment experiences?

2. Best-practices of assessment and information fostering learning

HOW—WHY—EXAMPLES

Prompting questions

- What do you consider best-practices with low-stake assessment?
- What made them a best practice?
- What were the consequences of this assessment?
- How did you feel about these consequences?

3. Room for not-discussed themes



CHAPTER 5

Setting decision-making criteria: is medical education ready for shared decision making?

Published as:

Schut, S., & Driessen, E. (2019) Setting decision-making criteria: is medical education ready for shared decision making? *Medical Education*, 53 (4). 324-6.

<https://doi.org/10.1111/medu.13826>

Setting decision-making criteria: is medical education ready for shared decision making?

Human decision making is prone to bias, fallibility and irrationality.¹ In high-stakes accountability systems, such as health care and education, this can be challenging to deal with. The quality of patient care and of competency-based education and assessment depends on the collaborative approach of multiple experts making numerous judgements while dealing with ill-structured problems. Both contexts are characterised by challenging, high-stakes work demands and by the crucial roles played by frontline professionals, both of which place substantial pressure on the quality of human decision making.² How should we deal with the limitations in our abilities to make and improve these high-stakes, complex decisions?

Heuristics offer some guidance and support when dealing with complex decision making in ill-structured settings. They are described as mental shortcuts or interrelated sets of principles or guidelines that are used to guide the process of problem solving.³ By contrast with algorithms (i.e., step-by-step prescriptions for achieving particular goals that, when used properly, promise a guaranteed solution to the problem), heuristics are problem-solving strategies that may lead to solutions. An example of heuristics is analogical thinking, in which one limits the search for solutions to situations that are similar to that at hand. Feufel and Flach⁴ propose to boost the use of heuristics in medical education to improve the process of clinical decision making. We echo the call made by Feufel and Flach⁴ and think it is interesting to make an analogy to the decision-making process in the assessment context within medical education. As we have alluded to earlier, we think there are relevant similarities in expert judgements and the complexity and stakes of the decision-making process in the respective environments of clinicians and educators.

Feufel and Flach⁴ present two commonly used heuristics they have identified in clinical decision making in the emergency department: ‘Common Things’ (what is most common given the symptoms, medical history and current observations), and ‘Worst Cases’ (in which the differential diagnosis is focused on the potential consequences associated with the different symptoms rather than the likelihood of a disease). In medical education one of the most high-stakes, summative decisions an educator is required to make concerns whether or not a learner has ‘succeeded’. This is a decision in which an educator needs to discriminate between those who know, understand, are on track, are competent and are entrusted, and those who are not. Indicators of learners’ performance levels or competency development may be more easily interpreted by teachers or supervisors with the use of these heuristics: Common Things (what is most common or likely given the indicators of behaviour and performance), and Worst Cases (alarming indicators of potential consequences related to the student’s performance that focus the teacher’s support on the provision of remediation or more supervision, such as indications of lapses in professional behaviour). This could, according to Feufel and Flach⁴ improve human decision making in clinical situations and may also work in the context of assessment in medical education.

“In medical education one of the most high-stakes, summative decisions an educator is required to make concerns whether or not a learner has ‘succeeded’”

Because there is significant overlap in the symptoms associated with both Common Things and Worst Cases, errors will occur. Feufel and Flach⁴ refer to those errors as false alarms and misses. In the assessment literature, they are known as false positives and false negatives. Feufel and Flach⁴ argue that the primary cause of error in decision making is therefore not human weakness or irrationality, but clinical complexity. To improve the quality of the decision and to mitigate the likelihood of errors, the authors borrow a concept from signal detection theory, the so-called ‘decision criterion’. This reflects the criterion that is used to decide whether to focus patient management on a Common Thing or a Worst Case and is a key parameter with respect to the quality of the decision. What determines ‘good’ quality or a ‘satisfactory’ criterion will depend on domain-specific values and the potential consequences of an error. In summative assessment, the responsibility and power to make that decision (or to ‘set the decision criterion’) are, in most cases, within the exclusive domain of the assessor, institute or regulatory body – a practice that is rarely challenged. It reflects the underlying norms and values we hold to assessment practices and who we consider best placed to make those high-stakes decisions. Does this practice still reflect the domain-specific values in medical education and assessment? Let's consider the values and practices that currently apply in assessment in medical education.

“Does current assessment practice still reflect the domain-specific values in medical education?”

Given its emphasis on self-regulated learning and the aspiration to deliver lifelong learners to the health care system, medical education clearly seems to value the role and responsibility of the learner. Are these values mirrored in our assessment practices and more specifically in setting the decision criterion? The pivotal question here is: who should set the ‘satisfactory’ decision criterion to make high-stakes decisions in assessment within competency-based medical education? Isn't the unilateral setting of the decision criterion by the assessor at odds with the objectives of modern education and assessment models, which aim to determine who is a reflective, self-regulating and competent professional?

“Who should set the ‘satisfactory’ decision criterion to make high-stakes decisions in assessment within competency-based medical education?”

If we want learners to make difficult choices in an informed and reflective way, such as in the complex clinical decision-making process presented in the paper by Feufel and Flach⁴ we need to reconsider learners’ roles and involvement in setting the decision criterion. Involving the learner in setting the decision criterion that is used to determine competence or entrustability might benefit the decision-making process tremendously, not only in terms of improving the decision-making process, but also by increasing the acceptability of the decision and, more importantly, the meaningfulness of assessment practices in general. In a recent study, learners

CHAPTER 5

argued that opportunities to influence and control the process of decision making in assessment stimulate a sense of agency and facilitate a shared responsibility for their learning and assessment experience. Moreover, learner agency in assessment enables the potential to use assessment as a learning opportunity for self-regulated learning.⁵

“Involving the learner in setting the decision criterion that is used to determine competence or entrustability might benefit the decision-making process tremendously.”

We are not arguing that learners should set their own assessment criteria completely, and neither are we ignorant of the challenges and tensions of self-assessment practices. We do think it is time to transform the common assessment model – in which the role of the student is passive – into an assessment model in which both assessor and student contribute to the decision-making process.

REFERENCES

1. Kahneman D, Slovic P, Tversky A. Judgment under uncertainty: heuristics and biases. Cambridge; New York: Cambridge University Press; 1982.
2. Edmondson AC, Higgins M, Singer S, Weiner J. Understanding Psychological Safety in Health Care and Education Organizations: A Comparative Perspective. *Res Hum Dev.* 2016;13(1):65-83.
3. Reigeluth C. Instructional-design theories and models Mahwah, NJ: Lawrence Erlbaum Associates; 1999.
4. Feufel MA, Flach JM. Medical education should teach heuristics rather than train them away. *Med Educ.* 2019;53(4):334-44.
5. 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):654-63.



CHAPTER 6

Meaningful mapping of remediation in longitudinal and developmental assessment models

Published as:

Heeneman, S., & Schut, S. (2020). Meaningful mapping of remediation in longitudinal and developmental assessment models. *Medical Education*, 54 (10), 866-8.

<https://doi.org/10.1111/medu.14258>

Meaningful mapping of remediation in longitudinal and developmental assessment models

Comprehensive assessment practices have the potential to stimulate and shape learning. St-Onge et al.,¹ explore the current evidence pertaining to the implementation of developmental progress assessment (DPA), which is defined as assessments mapped on developmental pathways and is used to provide guidance to trainees and faculty members. In doing so, they offer important insights on longitudinal and developmental pathways and the role of assessment therein.

‘... developmental progress assessment (DPA), which is defined as assessments mapped on developmental pathways and is used to provide guidance to trainees and faculty members.’

In DPA, the mapping of assessments to pathways is focused on the learner's progression and has the potential to facilitate the development of competence in ways that could promote students' empowerment.¹ At the same time, St-Onge et al.,¹ present the identification of trainees in need of remediation as one of the positive outcomes of the implementation of DPA. This juxtaposition of student empowerment with programme-required remediation creates an interesting quandary. What is the role of remediation when assessments are intended to be mapped on developmental pathways that encourage students taking more control over their progression?

‘This juxtaposition of student empowerment with programme-required remediation creates an interesting quandary.’

The concept of remediation in medical education is widely debated. In a thematic review, Cleland et al.,² concluded that most remediation interventions focus on improving performance to pass a resit of an assessment instead of emphasising the development of learning. In addition, the analysis showed that what worked in remediation practices (let alone why) could not be delineated yet. In traditional assessment settings, resits are often pre-scheduled, and mandatory. If DPA follows a predetermined pathway with mapped assessments, remedial assessments could become traditional resits. This may require trainees' pathways to be stalled until they have passed the next mapped assessment, which seems contrary to the objectives of developmental pathways as presented by St-Onge et al.,¹ Developmental progress assessment models explicate expectations and provide a roadmap for each training stage to guide individual learners' development towards independent practice. So how can remediation be meaningfully integrated as part of longitudinal and developmental assessment models?

‘... remedial action can be integrated as part of a personalised pathway in DPA, helping the learner maintain a certain ownership ...’

Ellaway, Chou and Kalet³ offer an interesting model, ‘situated remediation,’ that offers greater compatibility with assessment models focused on progression and development. In

situated remediation, two intertwined systems are presented, one with a focus on success and completion, the other with a focus on failure and exclusion; remediation is considered to be a bridge or interface between the two systems. In the success system, students could perform below standard, but the teaching and assessment activities are planned or mapped to enable the learner to meet the expected level of performance. Ellaway, Chou and Kalet³ define this a ‘remedial action’ that is supportive, informal and short-term. In contrast to failure-focused assessment, this model ensures that remedial action can be integrated as part of a personalised pathway in DPA, helping the learner maintain a certain ownership (i.e., enabling the learner to map remedial activity as part of his or her developmental pathway with a focus on progression).

When diagnostic information from a DPA identifies a learner in need of remediation who cannot be expected or assumed capable of leading such him or herself,¹ the model of Ellaway, Chou and Kalet³ moves the learner to the remediation zone, where specific goals are defined along with a timeline for completion and mapped primarily by staff rather than the learner. The key distinction relative to some other models is that Ellaway, Chou and Kalet³ clearly propose structured remediation as part of medical education systems, not as afterthoughts or an ‘outsider’ activity. This fits with the purpose of DPA by virtue of allowing a clear developmental map to be created for assessment that identifies pathways able to provide guidance to trainees.¹ The existence of a failure subsystem, in which the learner is suspended or required to retake a component of the programme or, if unsuccessful, may be excluded, fits the purpose of integrating and mapping remediation activity as part of longitudinal and developmental assessment models that take idiosyncratic needs into account, rather than a standalone that forces a single strategy on every trainee who struggles.

‘... structured remediation as part of medical education systems, not as afterthoughts or an ‘outsider’ activity’

Identification and support of learners in need of remediation using assessments meaningfully mapped on longitudinal and developmental pathways resonates with the principles and models for systems of assessment.^{4, 5} In a system of assessment, each assessment point or feedback opportunity is meaningful and optimised for learning. The need to work on certain improvement points (e.g., a lack of certain knowledge and skills or the need for feedback on professional competencies) is based on the diagnostic information that should be generated by DPA, after a process of reflection or self-assessment by the learner, and thereby becomes personalised. Doing so requires that the curriculum and the mapping of assessments in the model of DPA must have sufficient flexibility to provide ownership for the learner to facilitate planning of remedial actions that will help them improve.

‘... integrating and mapping remediation activity as part of longitudinal and developmental assessment models that take idiosyncratic needs into account...’

CHAPTER 6

One example that illustrates this principle is progress testing - a comprehensive test tailored to the learning objectives learners should achieve by the end of the curriculum that is typically administered several times per year.^{6, 7} Deliberately mapping progress tests to achieve a longitudinal and developmental assessment model can allow students to actively use feedback generated by sequential tests to determine how their overall progress is taking place. Students have indicated that by using and reflecting on progress test feedback, they can actively work on certain improvement points and remediate during the normal curriculum, suggesting that students can map their learning, undertaking remedial action and navigating in a longitudinal and developmental way, without being stalled on their developmental path.⁸

The curriculum and the mapping of assessments in the model of DPA must have sufficient flexibility to provide ownership for the learner to facilitate planning of remedial actions that will help them improve

The scoping review by St-Onge et al.,¹ was valuable to call attention to the mapping and personalisation of DPA, stimulating many thoughts on what remediation entails and how it can be meaningfully integrated in longitudinal and developmental assessment models. These insights can be considered as important steppingstones for the further development of longitudinal and developmental assessment models.

REFERENCES

1. St-Onge C, Vachon Lachiver E, Langevin S, Boileau E, Bernier F, Thomas A. Lessons from the implementation of developmental progress assessment: A scoping review. *Med Educ.* 2020;54(10):878-87.
2. Cleland J, Leggett H, Sandars J, Costa MJ, Patel R, Moffat M. The remediation challenge: theoretical and methodological insights from a systematic review. *Med Educ.* 2013;47(3):242-51.
3. Ellaway RH, Chou CL, Kalet AL. Situating Remediation: Accommodating Success and Failure in Medical Education Systems. *Acad Med.* 2018;93(3):391-8.
4. Norcini J, Anderson MB, Bollela V, Burch V, Costa MJ, Duvivier R, et al. 2018 Consensus framework for good assessment. *Med Teach.* 2018;40(11):1102-9.
5. Van der Vleuten CPM, Schuwirth LWT, Driessen EW, Govaerts MJB, Heeneman S. Twelve Tips for programmatic assessment. *Med Teach.* 2015;37(7):641-6.
6. Van der Vleuten CPM, Verwijnen GM, Wijnen WHFW. Fifteen years of experience with progress testing in a problem-based learning curriculum. *Med Teach.* 2009;18(2):103-9.
7. Freeman A, Van der Vleuten C, Nouns Z, Ricketts C. Progress testing internationally. *Med Teach.* 2010;32(6):451-5.
8. Heeneman S, Schut S, Donkers J, Van der Vleuten C, Muijtjens A. Embedding of the progress test in an assessment program designed according to the principles of programmatic assessment. *Med Teach.* 2017;39(1):44-52.



CHAPTER 7

Where the rubber meets the road: An integrative review of programmatic assessment in health care professions education

Published as:

Schut, S., Maggio, L.A., Heeneman, S., Van Tartwijk, J., Van der Vleuten, C., & Driessen, E. (2021). Where the rubber meets the road - An integrative review of programmatic assessment in health care professions education. *Perspectives on Medical Education*, 10(1). 6-13.

<https://doi.org/10.1007/s40037-020-00625-w>

ABSTRACT

INTRODUCTION Programmatic assessment was introduced as an approach to design assessment programs with the aim to simultaneously optimise the decision-making and learning function of assessment. An integrative review was conducted to review and synthesise results from studies investigating programmatic assessment in health care professions education in practice.

METHODS The authors systematically searched PubMed, Web of Science, and ERIC to identify studies published since 2005 that reported empirical data on programmatic assessment. Characteristics of the included studies were extracted and synthesised, using descriptive statistics and thematic analysis.

RESULTS Twenty-seven studies were included, which used quantitative methods (n=10), qualitative methods (n=12) or mixed methods (n=5). Most studies were conducted in clinical settings (77.8%). Programmatic assessment was found to enable meaningful triangulation for robust decision-making and used as a catalyst for learning. However, several problems were identified, including overload in assessment information and the associated workload; counterproductive impact of using strict requirements and summative signals; lack of a shared understanding of the nature and purpose of programmatic assessment; and lack of supportive interpersonal relationships. Thematic analysis revealed that the success and challenges of programmatic assessment were best understood by the interplay between quantity and quality of assessments information, and the influence of social and personal aspects on assessment perceptions.

CONCLUSION Although some of the evidence may seem compelling to support the effectiveness of programmatic assessment in practice, tensions will emerge when simultaneously stimulating the development of competencies and assessing its result. The identified factors and inferred strategies provide guidance for navigating these tensions.

INTRODUCTION

Programmatic assessment is a specific approach to designing assessment programs with the aim to simultaneously optimise the decision-making and learning function of assessment. Although all educational programs using a variety of assessments can be identified as assessment or evaluation programmes, programmatic assessment refers to a systematic approach to purposefully integrate a series of individual, dual-purposes measurements.¹ Two principles are considered unique and distinctive for programmatic assessment; the principle of proportionality and the principle meaningful triangulation,² both defined further below.

Compared to the traditional formative or summative purpose of assessment, programmatic assessment implies an assessment continuum which ranges from individual low-stakes assessments to high-stakes decisions.³ This means that all types of formal and informal assessments and feedback are low-stakes and focused on providing progress information to support learning in all competency domains. In programmatic assessment, information about learners' competence and progress is purposively and continually collected and analysed. There is an emphasis on longitudinal assessment of the learner, which creates a continual flow of information. High-stakes decisions are based on expert judgments of students' progress, which requires interpretation of the combination of results of a variety of assessment methods.^{1, 3, 4} The principle of proportionality refers to the association between the stakes of an assessment decision and the richness of the information on which that decision is based. Furthermore, contrasted to the traditional one-instrument-to-one-competency domain assessment programme, all competency domains are informed by various information sources,^{1, 4} which is referred to as the principle of meaningful triangulation.

In summary, programmatic assessment involves the careful selecting and combining of a variety of assessment methods and activities, which are meaningfully embedded in educational design, in order to promote learning as well as to obtain 'a whole picture' of learners' performance to inform high-stakes decisions.^{1, 3, 4}

There has been great interest in programmatic assessment, as it theoretically aligns with the goals and curricula of competency-based medical education.^{5, 6} Programmes ranging from undergraduate to postgraduate medical training are rapidly implementing the approach.^{7, 8} However, few efforts have been undertaken to examine the current state of research on programmatic assessment in practice. In this integrative review, we synthesise the available research to understand if and how programmatic assessment supports learning as well as decision-making, in order to inform the development of theory and to guide future research and implementations.

METHODS

We conducted an integrative review of the literature. In this approach, literature on a topic is reviewed using a systematic process in order to synthesise data from studies that have used a variety of methodologies.⁹ An integrative review offers a flexible approach in developing a rich and detailed interpretation to gain a comprehensive representation and understanding of programmatic assessment. This process is systematic but subjective by nature.¹⁰ Therefore, our research team represented a diversity of backgrounds and expertise, ensuring robust and critical interpretation of the included studies. SS is a PhD candidate focusing on the educational consequences of assessment. LM is a researcher and information scientist with expertise in conducting knowledge syntheses. CvdV has extensive expertise in developing and investigating programmatic assessment. SH is a program director and experienced in implementing curricular innovations in health care professions education. ED and JvT are experienced researchers in assessment and teacher education, respectively. Our approach and findings are reported in accordance with the STORIES statement.¹¹

Search strategy

SS and LM co-designed the search strategies for PubMed, Web of Science, and ERIC. Combinations of keywords and controlled vocabulary terms were optimised for each database (see Appendix I for the complete search strategies). Initial searches were conducted on July 9, 2019. Searches were limited to citations published from 2005 to the present, as the seminal paper proposing a programmatic approach to assessment was published in 2005.¹ A follow-up search was conducted on December 8, 2019, to assess for new, relevant articles. To ensure comprehensiveness of the approach, bibliographies of included studies were manually hand searched for additional relevant cited and citing studies.

Eligibility Criteria

Two reviewers (SS and a research assistant) independently screened all titles and abstracts. Publications were included when the implementation of programmatic assessment followed the principles of meaningful triangulation and of proportionality as described in the Introduction. To be included, studies had to contain evidence of the impact of programmatic assessment in practice by collecting and analysing empirical data; however, at this point no judgement was made regarding the robustness of the data. Studies were excluded if they were not in English, published before 2005, did not include a health care professions population, or focused solely on the development of programmatic assessment or describing the local implementation process. Review articles, commentaries, and letters were excluded.

Data extraction and analysis

We developed a data extraction form, which was operationalised in Microsoft Excel. Descriptive and citation information was extracted, including but not limited to study setting, study aim, study design, study population, data collection methods, and summary of key findings. We characterised the implementation of programmatic assessment using the principles of meaningful triangulation and of proportionality as described in the Introduction. Using Kirkpatrick's Hierarchy,^{12, 13} we classified study outcomes on four levels: satisfaction/perception, learning outcomes, performance improvement, and patient/health outcomes. We summarised the data extraction results using descriptive statistics. To supplement these findings, we thematically analysed¹⁴ the results of each study to identify themes in the key findings and mechanisms that enabled or hindered programmatic assessment to operate in practice. Each study was independently analysed and coded using an open coding strategy. SS analysed and coded each study in conjunction with one other independent reviewer (ED, LM, SH, CvdV, or JvT). The two coinvestigators met after extracting data in each subset to review the coding process and define and agree on themes (i.e., patterns within the data) emerging from the process. In a process of constant comparison, SS collated all proposed themes, which the research team collectively discussed to create a comprehensive and shared understanding of each theme. Disagreements were discussed and resolved by the whole research team to either reach consensus or further refine the description of the proposed theme. All authors reviewed and agreed on the themes arising from the data analysis.

RESULTS

Based on our inclusion and eligibility criteria, 27 publications¹⁵⁻⁴¹ were analysed (see Appendix II for the literature search and selection process). We first present the general state of the literature, which is summarised in Table 1, and then the identified themes. The complete overview of studies is available as Appendix III.

General description of literature on programmatic assessment

The 27 studies used quantitative methods (n=10), qualitative methods (n=12) or mixed-method (n=5) approaches. All used a competency-based educational framework and often noted that this was the reason for implementing programmatic assessment. Twenty-one studies were conducted in workplace settings (i.e., the clinical phase of medical education), three within a pre-clinical setting, and three within both clinical and pre-clinical settings. Thirteen studies explored learners' perceptions (e.g., medical students, interns, residents), 12 explored teachers' perceptions (e.g., assessors, mentors, coordinators, preceptors, supervisors), and 13 studies used assessment data (e.g., number of mini-CEX, grades, portfolio judgements, or quality evaluation of portfolio judgements). The majority of the studies originated from either Europe, specifically the Netherlands (n=10), or North America (n=9). Based on Kirkpatrick's Hierarchy,^{12, 13} we classified 17 studies at level 1 (satisfaction), three at level 2 (learning outcomes), and one

study reported outcomes on both levels. No studies were observed at level 3 (performance improvement) or 4 (patient/health outcomes).

TABLE 1. Summary of study characteristics.

	CHARACTERISTICS	N (%)
Study design	Quantitative	12 (44.4%)
	Qualitative	10 (37%)
	Mixed methods	5 (18.5%)
Implementation location	The Netherlands	10 (37%)
	Canada	6 (22.2%)
	United States	3 (11.1%)
	Australia	1 (3.7%)
	United Kingdom	1 (3.7%)
	Iran	1 (3.7%)
	New Zealand	1 (3.7%)
	Multiple locations ^a	4 (14.8%)
	Setting	Clinical
Pre-clinical		3 (11.1%)
Both		3 (11.1%)
Data sources ^b	Learner perceptions	13 (36.1%)
	Teacher perceptions	11 (30.5%)
	Assessment data	12 (33.3%)
Kirkpatrick levels	Level 1	17 (62.9%)
	Level 2	3 (11.1%)
	Level 1 and Level 2	1 (3.7%)
	Level 3 / Level 4	0 (0%)
	Not applicable	6 (22.2%)

^a 'Multiple locations' refers here only to some combination of the countries listed here

^b Multiple data sources in a single study add to the total of 36 data sources indicated here (100%)

Thematic results

We identified three themes. In the first theme, the dual purpose of programmatic assessment, we present the synthesis of insights on the decision-making and learning function of programmatic assessment separately to provide a clearer understanding of each assessment purpose in a dual-purpose context. Data analysis revealed that the success and challenges of the dual purpose of

programmatic assessment were best understood by four distinct albeit interacting factors. These are presented in the second and third theme: the interplay between quantity and quality of assessments information, and the influence of social and personal aspects on assessment perceptions.

Theme 1: The integration of the decision-making and learning function of assessment

Fourteen studies reported that programmatic assessment generated sufficient information to enable meaningful triangulation and high-stakes, robust decision-making.^{15, 16, 18, 19, 24, 25, 29-32, 37-40} Arguments for this conclusion involved findings such as high levels of assessors' agreement^{24, 30, 32, 37}; the perceived fairness or acceptability by learners and teachers^{23, 24, 30}; satisfactory reliability estimates^{15, 23-25, 29, 31, 37}; the coherent nature of the program that ensured all competencies were considered^{17, 18, 21, 22, 28-30, 34, 38}; and early detection of struggling learners,^{15, 18, 22, 24, 32, 38} specifically of problematic progression on the domains of professionalism and communication, which went undetected prior to implementation.^{15, 24} Furthermore, some assessors reported that the emphasis on learners' self-assessment enabled insight into learners' understanding of feedback and reflective skills as well as allowing triangulation of assessment information.^{24, 25, 30, 32} Although the introduction of promotion or clinical competency committees was found to enable or improve high-stakes decision-making,^{19, 30, 40} learners thought their mentors would be more credible in making this decision than such committees¹⁶ and supervisors often opted to rely on their own observations to inform their high-stakes decisions.³⁵ Making a high-stakes decision depended on the quality of the available information,^{19, 35, 40} and this quality was often found to be poor or even problematic.^{17, 20-22, 27, 34, 35, 40}

Fifteen studies concluded that programmatic assessment was beneficial to and could be used as a catalyst for learning.^{16, 17, 20, 22, 23, 25, 27-29, 31, 33, 34, 38, 40, 41} This conclusion was based on the findings that programmatic assessment supported learners' self-assessment and development as lifelong learners^{16, 20, 21, 24, 25, 27-30, 32, 38, 40}; increased learners' ownership^{21, 28, 30, 34}; allowed for targeted areas of improvement^{16, 20, 22, 24, 27, 28, 33, 34, 38, 40}; and shifted learners' perception to assessment information as a learning opportunity.^{16, 22, 27, 28, 30, 38} Based on the analysis of assessment information, three studies concluded that the approach benefited all learners to maximise their learning, including the students that were initially falling behind.^{25, 31, 32} However, assessors raised concerns that programmatic assessment could be more challenging for learners with a non-native background due to the essential role of narratives,³⁰ and one study reported the influence of learners' performance level on their feedback-seeking behaviour.²⁶ Although learners perceiving assessment as a decision moment did not necessarily mean it was not beneficial for their learning, studies also showed assessment was commonly perceived as high-stakes or hampering the learning opportunity.^{17, 20-22, 33} Specifically, the use of summative signals, such as grades, numerical scales on assessment instruments, pass/fail decisions, the obligatory nature of remediation or the mandatory uptake of assessment information in a portfolio, led to problems in interpretation of the nature and purpose of programmatic assessment. Summative signals could hamper learning opportunities, lead to more competition between learners, and result in loss of valuable information.^{17, 20-22, 33-35, 40}

Theme 2: The delicate interplay between quantity and quality of assessment information

The quantity as well as the quality of assessment and use of rich and narrative feedback was essential to create learning opportunities as well as to ensure meaningful triangulation for high-stakes decision-making.^{17, 19-22, 27, 34, 35, 40} The use of multiple assessments lowered the perceived stakes of individual assessments^{22, 33}; guided better recall of past performance and stimulated users to monitor progress over time and focus on trends^{16, 17, 20, 21, 25, 27, 28, 33, 35, 38, 40}; and improved identification of strengths and weaknesses to facilitate tailored learning programs.^{15, 16, 18, 22, 24, 27, 28, 32, 33} The programmatic approach sparked emphasis on direct observation and increased feedback^{19-22, 28, 29, 34, 38} and enhanced dialogue on performance progress.^{20, 22, 28, 29, 34, 40}

Although teacher and learners considered the use of multiple assessments and the resulting documentation pivotal for analysis and follow-up,^{17, 20-22, 27, 33, 40} adverse effects were also reported. The increased number of assessments created a heavy workload for teachers and learners^{20-22, 24, 28, 29, 33, 34} and risked assessment becoming viewed as time-consuming rather than meaningful or as a mainly bureaucratic activity.^{20, 22, 24, 33, 38} In workplace-based settings, programmatic assessment was reported to negatively impact clinicians' workflow.^{17, 21, 28, 34} This also impacted learners, who described a reluctance to ask for feedback due to their awareness of teachers' workloads.^{17, 22, 28, 33, 35} Assessors felt disinclined to provide honest or critical feedback because they feared the impact of feedback on learners or the extra workload that was thought to follow.^{21, 28, 34, 39, 40} Moreover, teachers' willingness to provide feedback was negatively influenced when they had to commit feedback to writing, which could corrupt the feedback content.^{17, 20, 21, 28, 34} Both teachers and learners expressed a fear of 'gaming the system', for example by strategical case or assessor selection.^{20-22, 34} Furthermore, the use of multiple assessments sometimes led to a perceived decrease in quality: teachers' tendency to give high marks impeded the monitoring of longitudinal development,^{16, 17} and generic feedback or no variation in assessment information provided limited input for the analyses of learners' strength and weaknesses.^{17, 18, 20, 21, 28, 34} This perceived decrease influenced learners' uptake of assessment as a learning opportunity^{16, 17, 20, 33} and teachers opting to use their own observations to inform decision-making rather than relying on the formal system.³⁵ The source of feedback played an important role in perceptions of assessment quality, like teachers' credibility,^{20, 22, 35, 38, 41} and although peer feedback was more easily perceived as formative,^{17, 20} teachers' feedback was often favoured over feedback from peers.^{20, 41}

Theme 3: The influence of social and personal aspects on assessment perceptions

We found a safe and supportive social environment to be pivotal for learning in programmatic assessment.^{16, 21, 30, 33, 41} Teacher support was identified as a key condition in fostering such an environment.^{16, 17, 20-22, 24, 30, 32, 33, 35, 39-41} Teachers described their roles as providing guidance, support, and resources; monitoring learners' progress; and facilitating constructive discussions.^{16, 30, 35, 38-40} Their contribution to high-stakes decisions could then result in role conflicts.^{34, 35, 39, 40} For some teachers, the introduction of an independent progress committee solved these conflicts.^{30,}

³⁸ However, for others such committees introduced difficulties in accepting loss of control over final decisions or highlighted their lack of trust in the assessment system. ^{17, 21, 30, 35, 39, 40}

To build trusting interpersonal relationships, which were deemed necessary for assessment to stimulate learning, teachers and learners required frequent interaction and time spent together. ^{16, 17, 21, 33, 40, 41} Eight studies highlighted the importance of learner agency in teacher-learner relationships. ^{16, 20, 21, 30, 33, 34, 39, 41} Although there was evidence that learners experienced more agency in programmatic assessment^{28, 30}; others reported lack of control and ownership, which could jeopardise the learning potential of programmatic assessment. ^{20, 33, 38, 41}

Eight studies reported the need for a shared understanding between teachers and learners of the nature and purpose of programmatic assessment. ^{16, 19, 21, 29, 30, 32, 38, 40} Achieving such understanding required sufficient training and faculty support (e.g., instructions, guidelines, consensus meetings). ^{16, 17, 22, 29, 30, 34, 38, 40} This shared understanding was further impacted by teachers' and learners' personal values and belief systems^{21, 28, 34, 35, 39}; their previous assessment experience^{22, 30, 33, 40}; the level of their experience with programmatic assessment^{16, 17, 20-22, 28-30, 33, 34, 38, 40, 41}; and learners' confidence level, motivation or orientation towards learning.^{20, 21, 26, 28, 33, 34, 39} Teachers and learners had to gain trust in and understanding of the system. ^{16, 20, 22, 28, 29, 33, 38, 41} In six studies, the implementation of programmatic assessment resulted in a shift in the assessment culture; towards one in which daily feedback was normalised and in which learning and self-reflection could thrive. ^{19, 24, 28, 30, 34, 38}

DISCUSSION

We reviewed 27 studies to generate novel insights on programmatic assessment. At first glance, the studies in our review seem to paint a rather optimistic picture of programmatic assessment in practice. Studies reported that programmatic assessment generated sufficient information to enable robust decision-making and could be a catalyst for learning. However, closer inspection revealed several problems. These included an overload in assessment information and an associated excessive workload, the counterproductive impact of using strict requirements and summative signals, lack of a shared understanding of the nature and purpose of programmatic assessment, and lack of supportive interpersonal relationships. In this section, we consider our findings in a broader context of health care professions education and assessment. In Table 2, we offer some inferred strategies from the literature to inform future implementations of programmatic assessment.

The use of multiple low-stakes assessments played a vital role in improving high-stakes decision-making and learning, however, lowering the stakes of individual assessments does not exempt us from ensuring that these assessments meet high quality standards, at least not if we want individual assessments as well as the system as a whole to remain meaningful for learning and decision-making. Although there is some 'safety in numbers', there is a fine line between scarcity

and overload in the use of multiple assessments. To ensure quality, it seemed counterproductive to enforce top-down decisions and mandatory requirements on how much or what type of assessment information should be used for learning and decision-making. This approach likely results in tick-box activities and in both learners and teachers gaming or even corrupting the system. Assessment information must serve the information needs of learners as well as teachers and searching for a magic number of assessment tasks seems superfluous. However, when learners and teachers are able to negotiate what constitutes a meaningful evidence base for learning as well as decision-making in their own context, this allows for more ownership and engagement with a programmatic approach by both, which is a condition for its success.

Creating a safe and supportive learning environment is pivotal for unlocking the potential of programmatic assessment. In this endeavour, committing to prolonged and supportive interpersonal relationships seemed to be a requirement for both teacher and learners. In health care professions education, it is often inevitable that teachers and learners only work together for a brief period of time. This places certain limitations on building trusting teacher-learner relationships and can explain why the exchange of assessment information is ineffective.

No matter how well individual assessments or assessment programs are designed or intended, if teachers and learners do not understand or agree with their function or purpose, they will likely become trivialised. Programmatic assessment requires a transition from a teacher-directed approach to a learner-led assessment approach in which the development of expertise is emphasised. This change in roles and responsibilities for teachers as well as learners will almost inevitably cause uncertainty and involve resistance. Teachers and learners are more likely to support and invest in such a change if they subscribe to its educational value and are empowered to assume ownership of it. When it is clear how this change solves problems they might have, instead of creating new ones, teacher and learners might be more inclined to endorse this transition.⁴² Their understanding and commitment to this transition can be aided by their active involvement in the entire process, from the start of the decision to implement a programmatic assessment approach to its operationalization. This requires sufficient training, instructions, guidance, and support systems. Moreover, change takes time and requires strong leadership, the importance of which should not be underestimated.

Notably, the vast majority of the included studies were conducted in the clinical setting, in which work-place based assessment plays a major and vital role. This raises the question if a preclinical setting, one in which the lower levels of Millers' competency pyramid⁴³ are often more dominant in learning and assessment practice, is aligned with the underlying assumptions of programmatic assessment. Perhaps the approach is more compatible and therefore appealing for assessment in the clinical setting. This is an interesting question for future research. In a similar vein, we found that the majority of included articles originated in the Netherlands. This raises potential questions about the use of programmatic assessment outside of the Dutch context and provides a warrant for future researchers to conduct further investigation.

TABLE 2. Inferred strategies from the literature to improve the value and use of programmatic assessment.

Inferred strategy and exemplifying references
Build on creating a shared understanding of programmatic assessment by clearly introducing the nature and purpose, providing explanatory guidelines for individual assessments and how they are used in the system as a whole, and involving teachers and learners in the whole chain of the system. ^{16, 19, 21, 29, 30, 32, 38, 40}
Provide teachers and learners with feedback on the quality of provided assessment information and how their input contributes to the decision-making process. ^{17, 21, 24, 40}
Normalise daily feedback, observation, and follow-up, as well as reflection and continuous improvement. ^{19, 21, 22, 28, 34, 38}
Be cautious with mandatory requirements, being overly bureaucratic, and the use of summative signals in the design of programmatic assessment, ^{17, 20-22, 24, 28, 33-35, 40} but keep the approach flexible, fit for purpose and negotiable, specifically in relation to the information needs of different stakeholders and the realities of the educational context. ^{16, 17, 20, 21, 24, 28, 33, 34, 41}
Promote learner agency and the development of life-long learner capabilities by increasing learners' ownership over the assessment process. ^{20, 28, 30, 34, 41}
Address learners' and teachers' assessment beliefs and the implications of a learner-led assessment approach ^{21, 28, 34, 35, 39} and provide mentorship for novices within programmatic assessment; ^{16, 17, 20-22, 28-30, 33, 34, 38, 40} more experienced stakeholders can help with the transformation.
Invest in prolonged and trustworthy teacher-learner relationships to create a safe and supportive environment. ^{16, 17, 21, 33, 35, 39-41} Frameworks such as 'The Educational Alliance' model ⁴⁴ and the R2C2 model ⁴⁵ might be helpful in this respect.
Organise group discussions and ensure shared decision-making, these do not only ease teachers' individual assessment responsibilities but can also improve the assessment outcome. ^{19, 24, 30, 32, 34, 35, 40}
Invest in credibility and trustworthiness as quality concepts for stakeholders, the process, and the system. ^{21, 24, 34, 40} Norcini et al., ⁴⁶ offer a quality framework for assessment systems.
Ensure a supportive infrastructure (i.e., available time and resources, effective technology, and sufficient faculty development), while taking the realities of the educational context into account. ^{17, 21, 28, 34, 38, 40}
Offer leadership in times of change. Cultural change takes time and, although issues should be addressed quickly, programmatic assessment will not be implemented perfectly from the start. ³⁸

Limitations

This review must be considered in light of its limitations. Studies that did not present their assessment programme as programmatic assessment while still investigating the same principles, might have been missed due to the nature of our inclusion criteria. Despite our efforts to be comprehensive in our searches, it is possible that we inadvertently missed a paper even though we took steps such as hand searching the papers included to safeguard against this possibility. With respect to the holistic approach of meaningful triangulation, it is important to note that several studies in our review focused on a single assessment instrument, for example the mini-CEX,^{21, 22, 36} the progress test,²⁷ or a single competence like professionalism.^{25, 29} Additionally, our search strategy yielded a number of studies describing valuable lessons drawn from the local implementation process. Although these would provide valuable lessons for others aiming to implement programmatic assessment, they were outside the scope of this review.

CONCLUSION

This study adds to the literature by comprehensively collecting and reviewing studies that examined programmatic assessment in practice. Although some of the evidence in the literature may seem compelling to support the effectiveness of programmatic assessment in practice, tensions will emerge when simultaneously stimulating the development of competencies and assessing its result. The identified factors and inferred strategies provide guidance for navigating these tensions.

Acknowledgment: The authors wish to acknowledge and thank Veronika Duwel for her valuable work as research assistant in conducting this review.

REFERENCES

1. Van der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. *Med Educ.* 2005;39(3):309-17.
2. Uijtdehaage S, Schuwirth LW. Assuring the quality of programmatic assessment: Moving beyond psychometrics. *Perspect Med Educ.* 2018;7(6):350-1.
3. Van der Vleuten CP, Schuwirth LW, Driessen EW, Dijkstra J, Tigelaar D, Baartman LK, et al. A model for programmatic assessment fit for purpose. *Med Teach.* 2012;34(3):205-14.
4. Schuwirth LW, Van der Vleuten CP. Programmatic assessment: From assessment of learning to assessment for learning. *Med Teach.* 2011;33(6):478-85.
5. Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR. The role of assessment in competency-based medical education. *Med Teach.* 2010;32(8):676-82.
6. Norcini J, Holmboe E, Hawkins R. Evaluation challenges in the era of outcomes-based education. In: Holmboe E, Hawkins R, editors. *Practical Guide to the Evaluation of Clinical Competence*. 1e ed. Philadelphia, PA: Mosby/Elsevier; 2008. p. 1-9.
7. Eva KW, Bordage G, Campbell C, Galbraith R, Ginsburg S, Holmboe E, et al. Towards a program of assessment for health professionals: from training into practice. *Adv Health Sci Educ Theory Pract.* 2016;21(4):897-913.
8. Schuwirth L, Van der Vleuten C, Durning SJ. What programmatic assessment in medical education can learn from healthcare. *Perspect Med Educ.* 2017;6(4):211-5.
9. Whittemore R, Knaff K. The integrative review: updated methodology. *J Adv Nurs.* 2005;52(5):546-53.
10. Bearman M, Dawson P. Qualitative synthesis and systematic review in health professions education. *Med Educ.* 2013;47(3):252-60.
11. Gordon M, Gibbs T. STORIES statement: publication standards for healthcare education evidence synthesis. *BMC Med.* 2014;12:143.
12. Kirkpatrick D, Kirkpatrick J. *Transferring learning to behavior: using the four levels to improve performance*. San Francisco, California: Berrett-Koehler Publishers; 2005.
13. Curran VR, Fleet L. A review of evaluation outcomes of web-based continuing medical education. *Med Educ.* 2005;39(6):561-7.
14. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol.* 2006;3(2):77-101.
15. Wilkinson TJ, Tweed MJ, Egan TG, Ali AN, McKenzie JM, Moore M, 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;11:29.
16. Driessen EW, Van Tartwijk J, Govaerts M, Teunissen P, Van der Vleuten CP. The use of programmatic assessment in the clinical workplace: a Maastricht case report. *Med Teach.* 2012;34(3):226-31.
17. Bok HG, Teunissen PW, Favier RP, Rietbroek NJ, Theyse LF, Brommer H, et al. Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Med Educ.* 2013;13:123.
18. Hauff SR, Hopson LR, Losman E, Perry MA, Lypson ML, Fischer J, et al. Programmatic assessment of level 1 milestones in incoming interns. *Acad Emerg Med.* 2014;21(6):694-8.
19. Chan T, Sherbino J, Mc MAPC. The McMaster Modular Assessment Program (McMAP): A Theoretically Grounded Work-Based Assessment System for an Emergency Medicine Residency Program. *Acad Med.* 2015;90(7):900-5.

20. 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):487-98.
21. Bok HG, Jaarsma DA, Spruijt A, Van Beukelen P, Van Der Vleuten CP, Teunissen PW. Feedback-giving behaviour in performance evaluations during clinical clerkships. *Med Teach.* 2016;38(1):88-95.
22. Castanelli DJ, Jowsey T, Chen Y, Weller JM. Perceptions of purpose, value, and process of the mini-Clinical Evaluation Exercise in anesthesia training. *Can J Anaesth.* 2016;63(12):1345-56.
23. Imanipour M, Jalili M. Development of a comprehensive clinical performance assessment system for nursing students: A programmatic approach. *Jpn J Nurs Sci.* 2016;13(1):46-54.
24. O'Brien CL, Sanguino SM, Thomas JX, Green MM. Feasibility and Outcomes of Implementing a Portfolio Assessment System Alongside a Traditional Grading System. *Acad Med.* 2016;91(11):1554-60.
25. Peeters MJ, Vaidya VA. A Mixed-Methods Analysis in Assessing Students' Professional Development by Applying an Assessment for Learning Approach. *Am J Pharm Educ.* 2016;80(5):77.
26. de Jong LH, Favier RP, Van der Vleuten CPM, Bok HGJ. Students' motivation toward feedback-seeking in the clinical workplace. *Med Teach.* 2017;39(9):954-8.
27. Heeneman S, Schut S, Donkers J, Van der Vleuten C, Muijtjens A. Embedding of the progress test in an assessment program designed according to the principles of programmatic assessment. *Med Teach.* 2017;39(1):44-52.
28. Li SA, Sherbino J, Chan TM. McMaster Modular Assessment Program (McMAP) Through the Years: Residents' Experience With an Evolving Feedback Culture Over a 3-year Period. *AEM Educ Train.* 2017;1(1):5-14.
29. Zijlstra-Shaw S, Roberts T, Robinson PG. Evaluation of an assessment system for professionalism amongst dental students. *Eur J Dent Educ.* 2017;21(4):e89-e100.
30. Bacon R, Kellett J, Dart J, Knight-Agarwal C, Mete R, Ash S, et al. A Consensus Model: Shifting assessment practices in dietetics tertiary education. *Nutr Diet.* 2018;75(4):418-30.
31. Bok HGJ, de Jong LH, O'Neill T, Maxey C, Hecker KG. Validity evidence for programmatic assessment in competency-based education. *Perspect Med Educ.* 2018;7(6):362-72.
32. Ross S, Binczyk NM, Hamza DM, Schipper S, Humphries P, Nichols D, et al. Association of a Competency-Based Assessment System With Identification of and Support for Medical Residents in Difficulty. *JAMA network open.* 2018;1(7):e184581.
33. 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):654-63.
34. 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):434-44.
35. Castanelli DJ, Weller JM, Molloy E, Bearman M. Shadow systems in assessment: how supervisors make progress decisions in practice. *Adv Health Sci Educ Theory Pract.* 2020;25(1):131-47.
36. de Jong LH, Bok HGJ, Kremer WDJ, Van der Vleuten CPM. Programmatic assessment: Can we provide evidence for saturation of information? *Med Teach.* 2019;41(6):678-82.
37. Favier RP, Vernooij JCM, Jonker FH, Bok HGJ. Inter-Rater Reliability of Grading Undergraduate Portfolios in Veterinary Medical Education. *J Vet Med Educ.* 2019;46(4):415-22.
38. Griffiths J, Dalgarno N, Schultz K, Han H, Van Melle E. Competency-Based Medical Education implementation: Are we transforming the culture of assessment? *Med Teach.* 2019;41(7):811-8.
39. Meeuwissen SNE, Stalmeijer RE, Govaerts M. Multiple-role mentoring: mentors' conceptualisations, enactments and role conflicts. *Med Educ.* 2019;53(6):605-15.

40. Rich JV, Fostaty Young S, Donnelly C, Hall AK, Dagnone JD, Weersink K, et al. Competency-based education calls for programmatic assessment: But what does this look like in practice? *J Eval Clin Pract.* 2020;26(4):1087-95.
41. 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):441-56.
42. Hargreaves E. Assessment for learning? Thinking outside the (black) box. *Camb J Educ.* 2005;35(2):213-24.
43. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med.* 1990;65(9 Suppl):S63-7.

References cited in Table 2 only

44. Telio S, Ajjawi R, Regehr G. The "educational alliance" as a framework for reconceptualizing feedback in medical education. *Acad Med.* 2015;90(5):609-14.
45. Sargeant J, Lockyer J, Mann K, Holmboe E, Silver I, Armson H, et al. Facilitated Reflective Performance Feedback: Developing an Evidence- and Theory-Based Model That Builds Relationship, Explores Reactions and Content, and Coaches for Performance Change (R2C2). *Acad Med.* 2015;90(12):1698-706.
46. Norcini J, Anderson MB, Bollela V, Burch V, Costa MJ, Duvivier R, et al. 2018 Consensus framework for good assessment. *Med Teach.* 2018;40(11):1102-9.

Appendix I: Database search strategy

Search conducted on July 9th, 2019, and rerun on December 8th, 2019

PubMed

("educational measurement"[mesh] AND programmatic [All Fields])
OR ("educational measurement"[All Fields] AND programmatic [All Fields])
OR ("assessment for learning"[All Fields])
OR ("assessment of learning"[All Fields])
OR ("programmatic assessment"[All Fields])
OR ("programmatic assessments"[All Fields])
OR ("Education, Professional"[Mesh] AND "programmatic assessment")
OR ("Students, Health Occupations"[Mesh] AND "programmatic assessment")
OR (((medic* OR nurs* OR dent* OR pharm*) AND education) AND "programmatic assessment"))
OR ("Education, Professional"[Mesh] AND ("assessment system" OR "system of assessment" [title/abstract]))
OR ("Students, Health Occupations"[Mesh] AND ("assessment system" OR "system of assessment" [title/abstract]))
OR (((medic* OR nurs* OR dent* OR pharm*) AND education) AND ("assessment system" OR "system of assessment" [title/abstract]))

Results: 892 citations as of 7/9/2019; 947 citations as of 12/8/2019

ERIC

(program* AND ("assessment for learn*" OR "assessment of learn*")) OR ("programmatic assess*") OR ("educational measur*" AND programmatic) OR "systems of assess*"

Note: Limited to 2005:

Results: 497 citations as of 7/9/2019; 504 citations as of 12/8/2019

Web of Science

(program* AND ("assessment for learn*" OR "assessment of learn*")) OR ("programmatic assess*") OR ("educational measur*" AND programmatic) OR "systems of assess*"

Note: 363 citations as of 7/9/2019; 465 citations as of 12/8/2019

Total search result: 1752 citations as 7/9/2019; 1916 citation as of 12/8/2019

Appendix II. Literature search and selection process

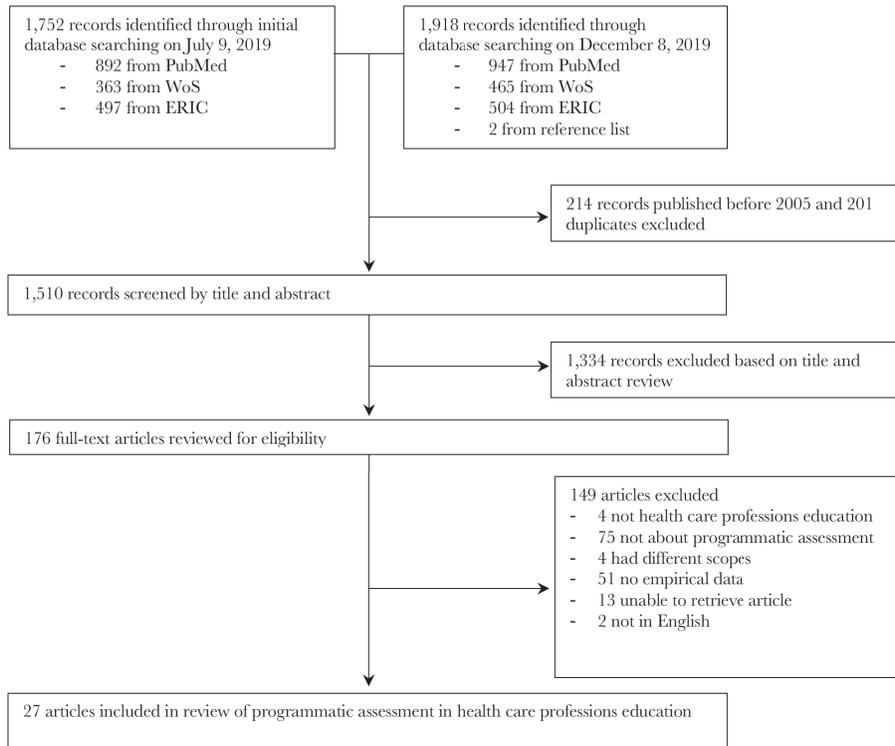


Figure legend. Literature search and study selection process in a review of programmatic assessment in health care professions educations, published 2005 to 2019.

Appendix III. Overview studies

Descriptions of 27 studies investigating programmatic assessment in practice in health care professions education, Published 2005 – 2019

Study ^{ref. no}	Location	Study design	Setting	Data source (n)	Focus of study	Level Kirkpatrick^a
Wilkinson et al, 2011 ¹⁵	New Zealand	Quantitative	Clinical	Assessment data (3539 & 826)	To describe and evaluate system to align learning and decision-making, and likelihood of passing compared with old system	2
Driessen et al, 2012 ¹⁶	The Netherlands	Mixed methods	Clinical	Learners' perceptions (670&19)	To gain more insight into learners' perceptions of the educational value and feasibility of assessment approach and robustness of final decision	1
Bok et al, 2013 ¹⁷	The Netherlands	Mixed methods	Clinical	Teachers' (8) & learners' (54&18) perceptions	To explore how theory interacts with educational practice	1
Hauff et al, 2014 ¹⁸	United States	Quantitative	Clinical	Assessment data (28)	To determine feasibility and results of programmatic assessment	NR
Chan et al, 2015 ¹⁹	Canada	Quantitative	Clinical	Evaluation data (50) - CCERR tool	To describe implementation of new system, report on initial experience, and compare before and after implementation	2
Heeneman et al, 2015 ²⁰	The Netherlands	Qualitative	Preclinical	Learners' perceptions (17)	To investigate the factors of implementation that students perceive as supporting or inhibiting learning	1
Bok et al, 2016 ²¹	The Netherlands	Qualitative	Clinical	Teachers' perceptions (14)	To explore factors influencing teachers feedback giving behaviour related to mini-CEX in system	1
Castanelli et al, 2016 ²²	Australia & New Zealand	Qualitative	Clinical	Teachers' (18) & learners' (17) perceptions	To explore learners' and teachers' experiences using mini-CEX in system	1

Study ^{ref. no}	Location	Study design	Setting	Data source (n)	Focus of study	Level Kirkpatrick^a
Imanipour et al, 2016 ²³	Iran	Quantitative	Clinical	Teachers' (8) & learners' (38) perceptions	To develop an assessment system, measure its validity, reliability, and educational impact	1
O'Brien et al, 2016 ²⁴	United States	Quantitative	Preclinical	Assessment data (156); learners' (79) perceptions	To evaluate the approach and summative portfolio judgements of portfolio review committee	1
Pecters et al, 2016 ²⁵	United States	Mixed methods	Preclinical	Assessment data (192)	To investigate triangulation of qualitative and quantitative assessments in system	2
de Jong, 2017 ²⁶	The Netherlands	Quantitative	Clinical	Assessment data (87)	To investigate if performance determines learners' feedback-seeking	NR
Heeneman et al, 2017 ²⁷	The Netherlands	Mixed methods	Pre-clinical & clinical	Assessment data (297); learners' perceptions (17)	To explore how a progress test embedded in programmatic approach affects test scores and learners' perceptions in comparative study	1 & 2
Li et al, 2017 ²⁸	Canada	Qualitative	Clinical	Learners' perceptions (26)	To explore and understand residents' experience perceptions of system	1
Zijlstra-Shaw et al, 2017 ²⁹	United Kingdom	Mixed methods	Clinical	Assessment data (81/66); teacher evaluations (19); learners' (9) perceptions	To determine the validity and reliability of assessment system and underlying framework	1
Bacon et al, 2018 ³⁰	Australia	Qualitative	Clinical	Teachers' (17) & learners' (29) perceptions	To explore stakeholders' experiences with a constructivist-interpretivist approach to decision-making in system	1

Study ^{ref. no}	Location	Study design	Setting	Data source (n)	Focus of study	Level Kirkpatrick^a
Bok et al, 2018 ³¹	The Netherlands	Quantitative	Clinical	Assessment data (327974)	To investigate use of programmatic assessment to track progression within and across competencies by providing validity evidence involving generalisation inferences	NR
Ross et al, 2018 ³²	Canada	Quantitative	Clinical	Assessment data (458)	To determine if programmatic assessment is associated with better identification and support for residents in difficulty compared with traditional assessment	NR
Schut et al, 2018 ³³	Canada, United States & The Netherlands	Qualitative	Pre-clinical & clinical	Learners' perceptions (26)	To explore learners' assessment perceptions and which factors influence perceptions	1
Acai et al, 2019 ³⁴	Canada	Qualitative	Clinical	Teachers' perceptions (16)	To explore teachers' experiences of the validity of competency judgments	1
Castanelli et al, 2019 ³⁵	Australia & New Zealand	Qualitative	Clinical	Teachers' perceptions (19)	To investigate supervisors experience using WBAs for summative assessment decisions	1
de Jong et al, 2019 ³⁶	The Netherlands	Quantitative	Clinical	Assessment data (352)	To provide evidence for saturation of information in portfolio assessment and the relation between final grade and number of datapoints	NR
Favier et al, 2019 ³⁷	The Netherlands	Quantitative	Clinical	Assessment data (574)	To assess inter-rater reliability of portfolio assessment	NR
Griffiths et al, 2019 ³⁸	Canada	Qualitative	Clinical	Teachers' perceptions (12&9)	To explore preceptors' perceptions of assessment culture and their changing roles in compared start and 3 years after implementation	1

Study ^{ref. no}	Location	Study design	Setting	Data source (n)	Focus of study	Level Kirkpatrick^a
Meeuwissen et al, 2019 ³⁹	The Netherlands	Qualitative	Clinical	Teachers' perceptions (12)	To explore mentor conceptualisation of dual role and role enactment in system	1
Rich et al, 2019 ⁴⁰	Canada	Qualitative	Clinical	Teachers' (12) & learners' perceptions (10)	To illustrate programmatic assessment model in action and identify strengths and challenges	1
Schut et al, 2019 ⁴¹	Canada, United States & The Netherlands	Qualitative	Pre-clinical & clinical	Learners' perceptions (26)	To explore the student-teacher relationship and the use of low-stake assessments	1

Abbreviations: mini-CEX, mini clinical examination, WBA: work-based assessment, and NR, not relevant.

^a Kirkpatrick's hierarchy included the following levels: level 1 (satisfaction/perception), level 2 (learning outcomes), level 3 (performance improvement), and level 4 (patient/health outcomes).



CHAPTER 8

Discussion

INTRODUCTION

In this dissertation, I have addressed the complex interplay between assessment and learning with a series of studies exploring the learning potential of a whole-system assessment approach in the context of competency-based medical education (CBME). The shift towards viewing assessment as a method to support and benefit learning and development instead of only for accountability might be inevitable,^{4,5} however, problems and concerns about the utility of assessment for learning remain persistent (**Chapter 1**). With the individual studies, I have aimed to contribute to the shifting discourse in medical education and to clarify the affordances of a whole-system approach coined as programmatic assessment. In this final chapter, I will review the main findings from different perspectives and discuss them in light of the overall aim and the extant literature. First, I will present and discuss the main findings of this dissertation from (1) a sociocultural perspective, and (2) an interpersonal perspective. Second, I will focus on the emergent themes arising from the overall findings, which all relate to the role of agency in assessment (1) agency and assessment objectives, (2) trustworthy assessment relationships, and (3) the role and importance of critical inquiry. Third, I will position the results within the wider educational context. Last, I will review the strengths and limitations of this dissertation, the implications for practice and future research, and finish with some concluding remarks.

Sociocultural perspective: conceptualising assessment stakes

The first part of this discussion considers the findings of the dissertation in relation to assumptions within a whole-system approach from a sociocultural perspective. This section will focus on the following two assumptions: (1) the concept of assessment as a continuum of stakes, and (2) lowering the stakes optimises the learning potential of programmatic assessment.

Assessment as a continuum of increasing stakes

As identified in the introduction, an important assumption in programmatic assessment is assessment as a continuum of increasing stakes. Each assessment is assumed to have a dual purpose (i.e., all assessment information is formatively used to support learning and summatively used for high-stakes decision-making). The study presented in **Chapter 2** explored how learners perceive assessment stakes and the concept of assessment as a continuum of increasing stakes. I found that learners refer to the same notion of assessment stakes as defined within the model of programmatic assessment, that is the consequences following an assessment. However, these consequences were not primarily considered as the proposed continuum in the programmatic assessment model, but rather as a dichotomy. Assessment comes with stakes (i.e., with consequences) or no stakes at all (i.e., no consequences), which related to whether or not their performance was observed and whether assessment evidence was documented or not (**Chapter 2** and **Chapter 7**). Learners have something to prove which consequently means that there is something at stake for them. Assessment as a continuum of stakes was recognised over time, meaning learners needed experience within the programme to get

familiar with and gain a shared understanding of the new assessment approach. Specifically, the use of multiple assessments combined with learners understanding of their coherence in relation to the high-stakes decision, lowered the perception of stakes involved in the individual assessment. However, in **Chapter 7**, adverse effects of the use of multiple assessments and the resulting documentation were also reported. The increased number of assessments created a heavy workload for teachers and learners⁶⁻¹³ and risked assessment becoming viewed as a time-consuming and mainly bureaucratic activity.^{6, 8, 9, 12, 14} Moreover, even with experience and understanding of the concept of assessment stakes, learners' conceptualisation of assessment stakes appeared rather complex, encompassing more than the formal consequences following an assessment. Learners' conceptualisation of assessment as a continuum and the stakes involved was strongly related to their perceived ability to act, control and make choices and decisions within the assessment environment, which was interpreted as learner agency (**Chapter 2**). The stakes of assessment therefore do not increase or decrease solely by detaching or changing the summative consequences of individual assessments, but shift more likely based on learners' sense of agency and their capacity to exercise that agency within the assessment process. Teachers' conceptualisation of assessment as a continuum was primarily influenced by the formal consequences following an assessment and focused more on the purpose of assessment (**Chapter 3**). Assessments on the low stakes end of the continuum had a clear purpose to stimulate and facilitate learning. However, likewise to learners' understanding of assessment stakes, the results showed that teachers' conceptualisations encompassed more than solely a focus on minimal consequences or the formative use of these assessments. According to teachers, the purpose of low-stakes assessments was to prepare learners for high-stakes assessments or future practice. This assessment conceptualisation strongly influenced how teachers facilitated learning: teachers thought a more directive approach was required to *ensure* learners were 'properly prepared' based on personal belief systems. Revealingly, teachers' conceptualisation of assessment stakes also had a controlling tone because assessment information concerning their learners was thought to mirror teachers' own performance.

The empirical studies presented in **Chapter 2** and **Chapter 3** showed that stakes are involved for both learners and teachers and that they are not as straightforward as the formal consequences following a single assessment. Furthermore, it became apparent that the vocabulary of an assessment continuum of increasing stakes refers primarily to the summative use of assessment information and implies a decision-making perspective on assessment. This might not be surprising given assessment's measurement problem from which the proposition of programmatic assessment departed (**Chapter 1**). In relation to this decision-making function, the knowledge synthesis in **Chapter 7** presented optimistic results. Studies evaluating local implementations of programmatic assessment concluded that the decision-making function of programmatic assessment was deemed robust, and in some cases even improved, due to high levels of assessors' agreement on the final high-stakes decision^{9, 15-17} or due to the perceived fairness or acceptability of the high-stake decision-making process.^{9, 15, 18} Both these claims were substantiated by the purposeful sampling strategy using a variety of multiple assessments, formats, methods, and assessors, and the documentation that followed. However, the aspirations with programmatic assessment go beyond improving

assessment's decision-making purpose. Although the insights concerning the conceptualisation of assessment as a continuum are helpful in relation to the further development of programmatic assessment in theory as well as in practice, the aim with low-stakes assessment is not so much to perceive assessment stakes as low, but to enable and support the learning potential of assessment. The question remains if by detaching the decision-making moment from the assessment moment the learning opportunity of assessment is truly unlocked.

Lowering the stakes of individual assessment optimises the learning potential of assessment

The other, yet related assumption underlying programmatic assessment from which I departed was that lowering the stakes of individual assessment would 'take the sting out of assessment'¹⁹,²⁰ and more likely enable learning opportunities for learners.^{1,21} As mentioned, I found learners' agency to be strongly related to learners' perception of assessment stakes, however, even more so to the perceived learning potential and value of assessment. A strong sense of learners' agency had the potential to engage learners more actively in the assessment process, positively influenced the perceived authenticity of assessment, learners' feedback receptivity and their willingness to learn from assessment. The value and importance of learners' agency for continuous development is supported and highlighted by many others e.g.²²⁻²⁸ and affects the ability or willingness to learn from assessment.^{22, 29, 30} Learners' agency was encouraged when the programme afforded learners to initiate their own assessments, to interact with and influence assessment outcomes, to show improvement, and when learners were able to select assessment and/or other relevant information to provide evidence of progress or achievement. An important design feature of programmatic assessment to enable such opportunities was the use of multiple assessments within a coherent structure. The use of multiple low-stakes assessments afforded learners and teachers to shift their focus from a 'snapshot' moment of learners' performance to a focus on longitudinal development in which they could monitor progress of performance over time and differentiate between trends and incidents in learners' performance. This change in focus was also an important finding of the knowledge synthesis presented in **Chapter 7**. Some authors even reported the implementation of programmatic assessment to result in a shift in their assessment culture: towards one in which daily feedback was normalised and in which learning and self-reflection could thrive.^{9, 10, 13-15, 31} In **Chapter 3**, teachers described how their assessment responsibility had changed from gatekeeper of performance achievement to one in which they conceived their responsibility as to support learners in discovering patterns in assessment information, to stimulate reflection, and to enable learners' improvement plans for reaching learning goals and perceived potential. The use of multiple low-stakes assessments afforded them to provide more honest and constructive feedback as consequences of a critical judgement were now limited to and focussed on improvement within the same learning cycle. Furthermore, receptivity to assessment information was deemed higher because learners and teachers could ground their interpretations and decisions on reoccurring information arising from various sources and formats. This type of receptivity is important if we aim for learners to be influenced by information and judgements generated by the assessment system.^{22, 26, 32-34}

Adversely, lack of opportunities for learners' agency led to more alienation from assessment and an overall high-stakes assessment experience (**Chapter 2** and **Chapter 4**). Specifically, the use of summative signals (e.g., grades, pass-fail decisions, mandatory demands) risked several problems, such as differences in understanding of the nature and purpose of programmatic assessment, competition amongst peers, untapped use of valuable information, and compliance behaviour. Grading low-stakes assessments and the use of norm-referenced strategies was found to encourage a focus on performance outcomes and competition rather than stimulating a focus on development and improvement. This adverse impact of grades and norm-referenced strategies is found and highlighted by many others in the field,^{e.g., 4, 22, 35-39} and reported to be detrimental for intrinsic motivation and task engagement.³⁹ The emphasis for learners becomes on the comparison and competition with others and their own ranking which takes away from learning and collaboration. This was found more often with assessments in the domain of knowledge and the use of standardisation, predetermined objectives and answer keys or criteria lists, which all provided little opportunity for learners' agency. According to learners and teachers, narrative feedback information was a more appropriate consequence of assessment in a whole-system approach, a finding which was additionally highlighted in the knowledge synthesis in **Chapter 7**. In order to stimulate and guide future learning, assessment information should focus on what learners can take away from an assessment and how this could impact future learning and development, instead of limiting the information to what is achieved at the specific task that moment in time (i.e., retrospective information).⁴⁰ Low-stakes assessment ideally includes information learners can use for further development (i.e., prospective information)⁴⁰ and empowers learners to demonstrate how they have used previous assessment information (**Chapter 2**). The implementation of the so-called post-assessment process or the follow-up activities⁴¹ is therefore an essential component when aspiring to stimulate the learning potential of low-stakes assessments, which in practice was not always the case (**Chapter 7**). Certain curriculum designs choices, often limited in flexibility and adaptivity, challenge opportunities for learners' longitudinal development, causing some programmes to opt for resits because certain themes or learning objectives are 'just not coming back' within the structure of the programme. In **Chapter 6** I have questioned the role of resits in programmatic assessment and offered a more aligned consideration of remediation which is integrated with learners' own educational pathways. The identification and support of learners in need of remediation using assessments meaningfully mapped on longitudinal and developmental pathways resonates with the principles a whole-system approach to assessment.^{3, 42} The need to work on certain improvement points is based on the diagnostic information generated by the assessment system so that, after a process of reflection and self-assessment by the learner, remediation can become personalised. Doing so requires the programme to have sufficient flexibility to provide ownership for the learner and to facilitate planning of remedial actions that will help learners improve. Moreover, as the vocabulary of programmatic assessment's continuum refers to the summative use of assessment information, we might need to consider introducing a narrative emphasising the formative use of assessment information, for example high-stakes learning opportunities instead of low-stakes assessments.

An interpersonal perspective: stakes in assessment relationships

The findings above highlight the important role and influence of learner agency in utilising assessment for learning. However, agency and such a notion of power never are static: what will be empowering for some, can be disempowering for others and could even be resisted by them.⁴³⁻⁴⁵ Empowerment of learners in programmatic assessment does not look the same for all learners involved, nor in all phases of medical education. Whether or not learners exercised control and acted upon the opportunities for agency depended on the interplay between experience and confidence, as well as their relationship with others (**Chapter 2**). Therefore, the following part of this discussion considers the findings from this dissertation on (1) a personal level, which considers the findings inherent to the learner or teacher, and (2) from an interpersonal perspective, which considers how learners perceive the interactions and relationships they engage in with their teachers.

Personal tendencies, past experiences and the influence of uncertainty

In **Chapter 2** previous experiences was identified as one of the key factors influencing learners' assessment perceptions. Most learners were used to defining their academic success based on their assessment performance. Their focus on a high grade-point average was deemed necessary in pursuing their ambition to become a health care professional, referring to the selection and admission process to enter a medical school or residency programme. Although learners generally considered themselves particularly successful in assessments (i.e., minimal experiences with failing), previous assessment experiences were primarily associated with high-stakes summative assessment environments, a high sense of dependence, pressure to perform well, the need to distinguish oneself based on assessment evidence, and with uncertainties, insecurities, anxieties, and fear of failing. Such associations and feelings had a strong and persistent influence on learners' assessment perceptions in the new system (**Chapter 2**). In the assessment literature, the influence of these factors is likewise recognised and identified as explaining the negative impact on the aspirations of using assessment for learning.^{e.g.,22, 46-49} In the context of assessment, feelings of uncertainty and insecurity were found undesirable for learners due to the perceived dependence on assessment performance. First-time experiences, for example with new assessment formats, assessors, or assessments in new settings, were unanimously perceived as high-stakes and thus require additional guidance and support. The influence of personal belief systems and previous assessment experience was also found with teachers as described in **Chapter 3** and **Chapter 7**. Teachers new in programmatic assessment voiced uncertainties about their understanding and proficiency with programme demands or lack of trust in the effectiveness of the assessment system. Like the findings in **Chapter 2**, these feelings led to a stronger desire for control over assessment processes and was further explained by teachers' high sense of responsibility to ensure and control quality, whether this was in relation to achievement of prerequisites that were thought necessary or in relation to patient safety when teaching in a clinical context.

Feelings of insecurity and uncertainty were considered undesirable for both learners and teachers in the context of assessment and explained the high-stakes perception of first-time

experiences, the initial dichotomy of assessment stakes instead of the envisioned continuum, and a desire for agency in assessment. However, in relation to the purpose of CBME and the characteristics of professional performance in complex high-stake settings, uncertainty might be inevitable. Both learners and teachers had to gain experience, confidence and trust in the meaning and consequences of a whole-system approach before they could shift their focus to a more learning-oriented perception of assessment. All studies showed prolonged and trustworthy teacher-learner relationships played a pivotal role in creating a safe and supportive assessment environment in which feelings of insecurity and uncertainty could be overcome.

Teacher-learner relationships in assessment

In **Chapter 4** I turned to interpersonal theory to explore and analyse learners' reflections on the interactions and relationships they engaged in with their teachers in the context of assessment. Given the importance of both learner and teacher agency in assessment described above, I was interested in how agency was negotiated within teacher-learner assessment relationships and how different types of interactions and relationships influenced assessment perceptions. Two important sources of teacher agency were identified: positional agency and expert agency. The impact of positional agency (i.e., agency that stems from holding certain authority positions) showed a prevailing focus on teachers' authority to validate and legitimise assessment information and judgement, even when learners did not consider their input meaningful or relevant for their learning objectives. Teachers' expert agency was found to increase assessment receptivity, to use assessment as a learning opportunity, and legitimised the influence of teachers on learner agency. Learners seek out what they consider relevant expertise and have a lot to gain from their teachers in relation to their expertise development. In that sense, desired expertise is at stake, which raises the learners' perception of assessment stakes. High levels of teacher communion had a positive impact on learners' perception of the assessment relationship as safe, trusting, and personal. Teachers with positional and/or expert agency who do not show dominance and who engaged in personal, committed relationships (i.e., relationships characterised by a high level of communion), were found most likely able to create safe assessment environments that afforded learners a sense of agency and a believed ability to exercise such agency. These teachers were described as having a genuine interest in learners and their aspirations, encouraging learners to share learning needs and providing guidance or direction when needed. Learners felt they could openly discuss their insecurities and uncertainties and invest in their improvement plans without fear of punitive consequences. They felt supported and could benefit from teachers' expertise without losing ownership over their learning process. These teachers were described as learners' advocates and seemed able, over time and based on trust, to 'transform'⁵⁰ some of their agency to learners.

Importantly, the introduction of an independent progress committee invited a third party to the teacher-learner assessment relationship. This afforded teachers to deal with a potential conflict between their supportive and their judgemental responsibilities more easily while still preserving a productive relationship with learners in an assessment context (**Chapter 3** and **Chapter 7**). Finding the right balance between supportive and judgemental responsibilities is a challenge

for teachers, which was already identified in the context of portfolio learning and assessment.⁵¹ Van Tartwijk and Driessen⁵¹ proposed a number of scenarios, of which their appropriateness depends on the educational context and the level of learners' experiences. Additionally, this dissertation found empirical support that the introduction of a third-party to the assessment relationship affords support, expertise and, more importantly, a safety net for both teachers and learners involved in programmatic assessment. Promotion and failure of learners becomes a shared responsibility and learners' careers do not rest on decisions made by individuals or on limited snapshots. This seems to release some of the emotional burden teachers can experience with assessment and allows them to provide feedback and raise concerns while preserving the benefits of prolonged engagement.^{37, 52-55}

Adversely, relationships that were characterised by an unjustified unequal power balance and/or a low level of communion, significantly increased the stakes involved in assessment. When teachers made their own agency prominent by behaving dominantly, learners felt more subordinate to their teachers, which constrained their own sense of agency, their believed ability to exercise such agency, and could hinder assessment being used for learning. Despite highly valuing learners' independence and self-regulation abilities, teachers were unlikely to lessen their control and direction over the assessment process and allow more learner agency, a finding presented in **Chapter 3** and discussed above. Power that stems from their positional agency. The prevailing influence of positional agency highlights learners' dependency relationship in assessment (**Chapter 2** and **Chapter 4**). Even when the teacher was perceived as unsupportive or lacked interest in learners' development, learners still looked for validation and approval and expressed that, '*at the end of the day*' is it the decision or judgement of the teacher that mattered. Learners remained focussed on teachers' judgement thereby maintaining the traditional and hierarchical nature and location of power in the assessment relationship,^{56, 57} constraining themselves from the exercise of their agency. In **Chapter 5** I have further questioned learners' dependency in assessment' decision-making process and whether the aspiration to equip learners for ongoing learning after formal education is truly mirrored in a unilateral decision-making process by teachers or the institute. Following the results of this dissertation, as well as the social-interpretivist paradigm in which I have positioned this dissertation, leads to the notion that teachers will view learners progress and performance through changing filters of values, beliefs, personal attraction or aversion, and mood. To then impose a unitary view of near-objectivity on the assessment process, is to require the learner to conform to the reality of the assessor.⁴⁵ If we want learners to make difficult choices in an informed and reflective way, learners' roles and involvement in assessment's decision-making process requires reconsideration.

Emergent themes

By discussing the findings from a sociocultural and an interpersonal perspective, I have aimed to clarify the affordances of programmatic assessment to support learning. These were: the use of multiple assessments in a coherent manner, characteristics of assessment information for learning and the role of follow-up activities, prolonged teacher-learner relationships, adaptive

guidance and support, and the introduction of an independent assessment committee. The findings in the individual studies all call for learner agency in the assessment process. As I reflected on these studies together, I came to see some overarching themes in relation to agency that require thoughtful consideration. These are agency and assessment objectives, trustworthy assessment relationships, and the role and importance of critical inquiry.

Agency and assessment objectives

The focus for the summative use of assessment information within programmatic assessment is on collecting and aggregating assessment information in such a way that the sum is more meaningful than its parts. One of the metaphors often used is that programmatic assessment thereby allows ‘to obtain a whole picture’ of students’ competence.^{1,58} In practice, programmes in many contexts need to comply to a national assessment or accreditation framework containing a set of predetermined outcomes in terms of knowledge, skills, and/or attitudes. Although the top-down determination of assessment objectives might be helpful ‘to obtain a whole picture’ of students’ competence, to ensure certain desired outcomes, and to effectively design coherent programmes, it risks constraining learner agency and could limit the opportunity for them ‘to construct a whole picture’ of their performance level. This certainly seemed to be the case with implementations that dictate which assessment information should be collected and documented rather than empowering learners to be accountable for their own learning (**Chapter 7**). The use of multiple assessments, while seemingly benign when conceived as lowering the stakes and creating multiple opportunities for learning, can soon create an stifling environment of constant observation.^{59, 60} Such an environment constrains agency rather than the liberating thoughts the model could theoretically have intended, and is more likely to result in instrumental or strategic approaches by learners and teachers, for instance tick-box activities, criteria compliance, and gaming or even corrupting the system (**Chapter 7**). Furthermore, continuous observation may be at odds with medical education’s value of progressive independence.^{25, 61} With dominant top-down, predetermined, and mandatory requirements, the assessment system can even become an instrument of hierarchical observation and normalising judgement: a controlling and directing instrument of constant surveillance and disciplining ‘its targets’ by judgement.^{43, 59, 60} Noteworthy, such a stifling process is most likely found in the domain of knowledge assessment (**Chapter 2-4**). Assessment of knowledge is often standardised and associated with the use of ‘objective’ scores and grades. Especially in the preclinical phases of medical education, assessment of the knowledge domain often plays a more dominant role in the overall assessment programme. According to both learners and teachers, standardised knowledge assessment is a different assessment ‘game’. It is in this domain that it is more likely to be faced with a prevailing hegemony in assessment that has faith in scientific notions of objectivity and reliability and that invest power in teachers’ authority.⁴⁵ Such a view makes the learner the ‘object’ of assessment practices,^{57, 62} affording little to no learner agency and could even alienate learners from their learning and assessment experience.^{60, 63} Implementing programmatic assessment thus requires thoughtful consideration of the role of predetermined outcomes and standardised assessments and how these are used in the programme and the

decision-making process, especially in preclinical phases of medical education. This is not to say they have no role to play. Providing learners with knowledge and skills and with forms of judgement that allow them to ‘do something’ and as such build on important knowledge traditions of the field, is considered one of the important purposes of education.⁶⁴ Furthermore, standardisation and predetermined objectives set by the programme certainly do not exclude the use of assessment information to support learning nor do they necessarily translate to a stifling environment. However, when replication of the field’s knowledge traditions becomes too dominant and predetermined objectives become too prescriptive, this does risk constraining learners and bringing their future learning into a known and narrow form. The construction of a whole picture of learners competence becomes a ‘painting by numbers’ experience, in which learners colourise their picture with predetermined and forced outcomes. As such, programmatic assessment could unwittingly set up a system that may incentivise learners to hide their weaknesses, emphasises the importance of commonalities, and drive to concretely define what is essential to being a health care professional. Proof of competence becomes the perceived goal of any assessment encounter, increasing the likelihood that learners will see this as a high-stakes hurdle. Finding a balance between the use of valuable high-quality ‘pixels’ that serve as meaningful steppingstones to learners’ expertise development in such a way that the sum is indeed more meaningful than its parts, and the use of such ‘pixels’ in which their sum is limited to and determined by the individual parts, is one of the more challenging balances that must be found in practice.

Trustworthy assessment relationships

This dissertation showed how pivotal trustworthy assessment relationships are to afford learners a safe and supportive assessment environment, which in turn lowered the perceived assessment stakes and more importantly, enabled the use of assessment for learning. It also showed how learners risk constraining their own agency by a primary focus on teachers’ positional agency and how difficult it is for teachers to empower learners in the context of assessment. Some feel the burden of proof rests solely on them to ensure, or even to guarantee, the quality of assessment outcomes. As assessment relationships are essentially unequal,^{57, 65, 66} there is a need to reduce this inequality⁵⁰ given the desired outcomes of CBME. Ideally, the teacher supports learners’ competency development and is able to achieve meaningful goals without trying to oppose or override learners’ agency. It requires frequent interaction and time spent together to build trust and it involves the transformation of teacher agency to learner agency over time and through development. For trustworthy relationships to develop, prolonged engagements are essential, which is not always possible in the context of medical education. Furthermore, goals of both teachers and learners should be compatible. In the context of medical education this should hold true if we state that ‘providing good quality health care’ is the ultimate goal, and both parties act in the best interest of the patient.⁶⁷ This is however, especially for undergraduate medical learners in the preclinical phase of their training, a long-term prospect, and several teachers and objectives in medical education do not have such a direct relation to patient care, such as those in the basic sciences (**Chapter 2-3**). Teachers might focus on preparing learners

for the next phase and learners might first focus on ‘surviving’ until the next phase, which could explain why some learners were seeking positional influence and felt more dependence towards teachers in the context of assessment. Furthermore, learners in undergraduate medical education might look for more interpersonal control from their teachers than medical trainees in postgraduate education, as the latter might have developed stronger self-regulation skills or at least have more confidence in their own abilities.

Although the value and importance of learners’ agency for continuous development is highlighted by many others in current views on learning,^{e.g. 4, 25, 30, 46, 49, 57} it remains a challenge in the context of assessment. The importance of learner agency in assessment, however, is not only a logical consequence of the social-interpretivist paradigm, but based on the findings of this dissertation, is key to enable the dual-purposes aspired with the introduction of programmatic assessment in CBME. Within a constructivist approach to learning and an assessment system that aims to purposefully integrate assessment and learning, it makes little sense to treat active involvement of learners in assessment any differently from agency in learning. The tension between trust and control and the adaptive need for guidance and support is another challenging balance that must be found in practice.

The role and importance of critical inquiry

High standards of competent achievement and ensuring quality is not only an important obligation of medical education to society-at-large,⁶⁷ but an important value both teachers and learners want to adhere to as was shown in **Chapter 2** and **Chapter 3**. The quantity as well as the quality of assessment and use of rich and narrative feedback is essential to create learning opportunities as well as to ensure meaningful triangulation for high-stakes decision-making (**Chapter 7**). Only focussing on learners’ agency in assessment when learners have no framework for challenging the ideas, criteria or standards put forward, creates difficulties for learners to challenging them and for teacher to trust and allow more learner agency in assessment. Furthermore, in a recent study by Watling et al.,²⁸ the exercise of agency in medical education was even identified as burdensome. More importantly, too strong a focus on learner agency can serve just as easily as a disciplinary process into this new discourse, which I have aimed to critique in the first place. Therefore, next to the empowerment of learners by a supportive assessment environment and trustworthy assessment relationships, the role of critical inquiry requires careful consideration. Here, critical inquiry concerns the interpretations given to assessment information generated by programme assessment and the narrative constructed based on this information by teachers and learners. This might provide an adequate ‘technology’ to balance and leverage the two identified tensions discussed above.

Critical inquiry refers to people developing a critical attitude to their own work through a reflective process which can lead to transformation of ideas and practice and is established by means of rational debate between different viewpoints that examine all assumptions and their consequences.^{44, 45} Boud and Molloy²⁷ advocate a curriculum approach to feedback where

learners have the opportunity to develop judgements about the quality, to engage in giving and receiving feedback through exchanges with others, and where such assessment information can be applied to future tasks and learning opportunities.²⁷ Eva et al.,²⁵ plea for ‘the translation of assessment experience into a personal inquiry-based learning strategy and the integration of the idea of using data to make sense of one’s experience and framing a plan for improvement’.^(p.902) Both of these calls highlight the importance of critical inquiry and afford more learner agency, and are empirically supported by the findings of this dissertation. In **Chapter 5** I have called to reconsider learners’ roles and involvement in assessment’s decision-making process. Empowering learners in assessment might not only benefit the quality of the decision-making process, but could also increase the acceptability of high-stakes decisions, the meaningfulness of the assessment process, and more importantly, better equip learners for future learning after formal graduation. This is not to say that learners should be solely responsible for assessment’s decision-making function or carry all the weight of this burden, nor to appear ignorant of the challenges and tensions with self-assessment. However, I do believe it is time to transform the traditional model in assessment in which the learner is ‘but a passive actor’ or even an ‘object’, solely dependent on the judgement of teachers, to one in which both teacher and learner contribute collaboratively to the decision-making process. Can we constructively shift the burden of proof and bear this together with learners and involve them in, not only the construction of their own whole picture, but also empower them to make informed decisions concerning their competency and expertise development? In other words, shouldn’t they be the main stakeholder and owner of any claim concerning their own competence? If we truly aspire learners to be committed to continuous professional development and ongoing advancement of their field after graduation,⁶⁸ they should have the opportunity to be heard and the means to resist power formations during formal education. Learners should be able position themselves in relation to assessment criteria and standards and should have the opportunity to question and negotiate such criteria and standards. This would enable not only the empowerment of learners in the assessment system and potentially withstands a primary focus on replication, but also has the potential to ensure the quality of those interpretations and for the whole to become greater than the sum of its parts.

Based on the findings in this dissertation as well as in consistency with the chosen worldview central in this dissertation, it is important to understand what teachers and, maybe more so, what learners bring to the assessment practice, which includes their expectations, orientations, assumptions, past experiences, emotions, and interpretations of the culture in which both the teacher and the learner are situated. In order to utilise the learning potential of programmatic assessment, I have argued that the assessment process requires a shift to assessment as a process of co-inquiry to negotiate and determine what counts as legitimate. This should be supported by trustworthy interpersonal relationships in which teachers transform their agency towards learners over time and through development, and in a context celebrating critical inquiry to ensure quality and to meet society’s expectations of the medical profession.⁶⁷

Positioning the results within the wider educational context

So far, the focus of this general discussion has been on the affordances of programmatic assessment within the vacuum of an educational programme. In that sense, this dissertation and my understanding of the relationship between assessment and learning refers to the ‘within-system’ double duty of assessment. However, educational programmes do not ‘exist’ in isolation, they are a part of learners’ overall trajectories or educational pathways.⁶⁹ The use of information derived within the system simultaneously involves an external system or rather the ‘between-systems’ use of assessment evidence, for example in the transition from high school to undergraduate education or from undergraduate medical education to postgraduate medical training. Programmes are part of a wider educational and assessment system and are influenced by a range of contextual, disciplinary, and ecological factors.⁷⁰ As I have highlighted and shown, assessment is a process giving rise to strong feelings, hugely influenced by past experiences, external licensing exams, and perceptions of future assessment demands. When assessment is seen as a social and value-laden activity, a claim I’ve tried to make in this dissertation and which I’ve underlined in the methods chosen to investigate the central problem, we should also consider the wider social and cultural context in which assessment operates.^{46, 62}

Assessment in higher education is surrounded by debates about academic standards, preparing students for employment, measuring quality, and providing incentives.⁴⁶ Issues at government and national decision level deal with the relationship between education and society. Broadfoot⁷¹ summarised the main four generic societal functions of assessment that mediate the relationship between education and society. First, assessment of competence, which is rooted in the history of craftsmanship. Second, assessment mediated through selection based on educational performance as regulating competition for scarce societal opportunities. Third, assessment for determining the content of what is taught, which is increasingly used as political tool to direct curriculum priorities in the educational system as a whole. Fourth, which is a significant development of more recent decades, assessment as a means of control through mechanism of accountability such as payment by results and league tables of comparative performance. As such, assessment has the power to shape both individual lives and the education system as a whole, and, as a policy tool, the power to drive local and national educational priorities.⁷¹ To understand the impact of such power, current conceptions of accountability have to be considered. In the last few decades, we have seen a remarkable rise in interest in the measurement of educational outcomes, also referred to as ‘the age of measurement’.⁷² According to Biesta⁷² the notion of accountability has shifted to a narrow, technical managerial approach to accountability, where measurement has become an aim in itself, and indicators of quality, like student evaluations and test results, become the definitions of quality. This manifestation can be understood by the interest in providing an ‘evidence-base’ for educational practice, which focusses on the measurement of educational outcomes and their correlation with educational ‘input’. This development played an important role in the idea that educational outcomes can and should be measured. The strong impact of documented assessment information on learners’ assessment perception found in this dissertation is an example of that what was intended to serve as

indicators of quality in learners' competency development becomes the definition of the quality of their competence. This manifestation can further be observed in prescriptive approaches by government agencies concerning outcome-based assessment in education, expressed as detailed, standardised definitions of learning outcomes, to define as precisely as possible the scope and content of learning outcomes and competencies and to be far more explicit and detailed about the criteria for assessment decisions than has previously been the case.⁶⁵ Assessment decisions are thought to require evidence and for some even to require guarantees. We seem to seek and value 'proof'. Although acknowledging that accountability is important and that public institutions should be and should want to be accountable for what they are doing, when discussions of assessment are dominated by these accountability purposes, considerations of the impact of assessment on learners, teachers and learning itself can be eclipsed.³³ Moreover, Biesta⁷² argues that accountability can become problematic when measurements begin to drive and even replace judgements about what good education or outcomes are. Assessment remains a burden of proof. Furthermore, in the words of Biesta,⁷² 'the danger here is that we end up valuing what is measured, rather than that we engage in measurement of what we value'.^(p.43) O'Neill⁷³ argues that current conceptions of accountability reveal what she has called 'a culture of suspicion' which super impose managerial targets on bureaucratic processes burdening and even paralysing those that have to comply. According to O'Neill,⁷³ huge efforts go into ensuring trustworthy performance and increasingly sophisticated technologies are deployed to prevent and detect breaches of trust. A recent example became apparent during COVID-19 in the tremendous effort and resources that went into online proctoring systems to enable the continuation of high-stakes assessments in higher education during the pandemic. With the introduction of online proctoring systems, higher education maintained the perceived necessity for surveillance despite the questionable costs, which include the emotional burden and pressure that these systems impose on students. Outcome-based educational systems like competency-based medical education are suspect to prescriptive outcome-based regimes⁶⁵ and to define professional responsibilities with ever greater accuracy.⁷³ Perhaps, however, O'Neill⁷³ argues, the culture of accountability and the increasing call and perceived need for transparency and surveillance systems, actually damages trust, rather than supporting it.⁷³ Careful consideration of what kind of assessment information is documented in formal systems is warranted, given the findings highlighted in this dissertation.

Honestly, I think many of us would admit to fundamental doubts about whether what we truly value in education can be measured as such. It is this doubt and insight that created momentum for programmatic assessment in the first place.⁵ Despite the evolving discourse in the medical education literature, which acknowledges the limitations of objectivity in assessment⁷⁴ and which values the social-cultural and interpretivist perspective on assessment,^{75, 76} it is the apparent objectivity of assessment that provides the legitimisation for its key role in life's chances.⁷¹ Airasian⁷⁷ already argued that the legitimacy of assessments derives not so much from empirical evidence of their probable effectiveness or the actual impact assessment has on teachers, learners, and the curriculum, but from perceptions they evoke and the symbol of order and control they represent. Specifically, high stakes, external assessments have much to do with exerting symbolic control

either over the education system or in favour of society,⁶⁷ more so in educational contexts with an absence of a national curriculum.⁶² That external or centralised assessment for accountability can have an adverse impact on teaching and learning as they cannot meet the information needs of teachers and learners in order to be meaningful was already highlighted by Stiggins⁷⁸. Not unimportantly, competition plays a major role in learners' overall assessment trajectories, specifically in medical education as they are burdened by selection due to, at least partially, scarce educational opportunities. As a consequence, the stakes in assessment are considerably high for learners aspiring a future as health care professional as assessment results currently determine their chances to get into medical education. The implementations of programmatic assessment that formed the stage in this dissertation showed that a shift to a learning culture was possible (within-system): the assessment approach augmented by trustworthy assessment relationships afforded a 'safety hub' within learners' overall assessment trajectories. If we take the potential of programmatic assessment shown in this dissertation seriously, I think we need to engage in the conversation of how assessment is used between educational programmes (between-system). Specifically due to the strong influence found of past experiences, external licensing exams, and perceptions of future assessment demands. If we truly value the construction of a meaningful and holistic narrative to inform judgement and high-stakes evaluation, and if we claim that a whole-system approach can overcome serious unintended consequences of assessment, should we not allow these holistic judgements to inform decisions between programmes instead of opting for high-stake selection processes or external licencing examinations? The only remaining morally justified method for regulating competition for scarce societal opportunities would then be lottery.

Strengths and limitations

In addition to the strengths and limitations that were identified in the individual studies, this dissertation as a holistic research project must also be considered in light of certain limitations.

Most of the included studies rest on the perceptions and experiences of participants' reality as they themselves were able to recollect and verbalises these. There may be differences between what participants report they believe and intend to do versus what they actually believe and do. Their memories might have faded or changed over time and be (partly) inaccurate. This is a limitation inherent in the approaches taken in these studies. As was outlined in the introduction, the ambition to capture an objective event or universal truth is not aligned with my worldview. Rather I consider the lived experience as offering valuable insights for theory development and understanding of the ultimate impact of assessment on learning. However, I do believe observational research and discourse analyses would be valuable for future research to provide additional insights on the complex relation between learning and assessment and the ambitions to move towards a learning culture. Furthermore, the findings related to the importance of participants' experiences, the role of interpersonal assessment relationships, and the influence of time and development in general. Future research using a longitudinal design could shed more light on the development of learners' perception of assessment stakes and of teacher-learner relationships in the context of programmatic assessment. I have attempted to compensate for

this by purposefully sampling participants in different phases of medical training ranging from preclinical undergraduate medical school to postgraduate residency training. Future research might be able to provide additional insights if the development of learners, teachers, and their interpersonal relationships would be further observed and analysed over time.

As assessment is a complex interaction of learner, task, teacher and context characteristics,⁴⁸ transferability of our findings to other contexts is challenging. I purposefully included multiple research settings in the empirical studies which strengthened my confidence in the interpretation and understanding of the underlying mechanisms and which I consider a strength of this dissertation compared to single institute studies. However, the settings were similar in the sense that they were all western settings. So it is not inconceivable that medical students, residents, or their teachers and supervisors from these settings hold different perspectives compared to their counterparts in other areas of the world, specifically due to the situated and cultural nature of the topic, and the influence of national frameworks, regulatory bodies or more general assessment regimes on educational institutes. I have attempted to augment transferability by providing transparency concerning the context and methods used to collect and analyse data. More importantly, this approach gave me the opportunity to give voice to learners' and teachers' experiences, which I consider the most important strength of this dissertation and its relevance and contribution to the literature. Furthermore, conducting a knowledge synthesis not only contributed to a broader methodologic expertise development as a researcher, but it also afforded a much broader overview of findings concerning programmatic assessment in practice and the lessons learned by others. The knowledge synthesis offered the opportunity to triangulate the findings of the empirical studies with the work and findings of others and strengthened my confidence in my own interpretations. I do, however, stress the importance of replicating the study designs in different contexts and learning cultures for further understanding and transferability. Furthermore, ambitions to use assessment for more than accountability transcend the medical education domain. It would be valuable to conduct similar research projects in other domains.

Implications for practice

Practical implications of this dissertation appertain to curriculum designers aspiring to combine assessment *of* learning with assessment *for* learning and to teacher professionalisation. In relation to the latter, I have argued it is worthwhile to consider how learners perceive teachers interpersonally when aiming to create learning opportunities with the use of assessment. The Interpersonal Circumplex⁷⁹ was found useful to understand the influence of teacher-learner relationships on learners' assessment perceptions and which types of interactions led learners to perceive assessment as a learning opportunity. This model could be used to create awareness amongst teachers on how they influence learners' assessment experiences and how they might alter these interactions and engage in effective relationships, as has been done in the domain of teacher education.⁸⁰ Furthermore, this dissertation signalled several tensions likely to emerge with a whole-system approach to assessment.

In **Chapter 7**, I have inferred strategies from the literature to leverage these tensions and to support the value and use of programmatic assessment in practice, which I recall below:

- Build on creating a shared understanding of programmatic assessment by clearly introducing the nature and purpose, providing explanatory guidelines for individual assessments and how they are used in the system as a whole, and involving teachers and learners in the whole chain of the system.
- Provide teachers and learners with feedback on the quality of provided assessment information and how their input contributes to the learning and decision-making process.
- Normalise daily feedback, observation, and follow-up, as well as reflection and continuous improvement.
- Be cautious with mandatory requirements, being overly bureaucratic, and the use of summative signals in the design of programmatic assessment, but keep the approach flexible, fit for purpose and negotiable, specifically in relation to the information needs of different stakeholders and the realities of the educational context.
- Promote learner agency and the development of life-long learner capabilities by increasing learners' ownership over the assessment process.
- Address learners' and teachers' assessment beliefs and the implications of a learner-led assessment approach and provide mentorship for novices within programmatic assessment; more experienced stakeholders can help with the transformation.
- Invest in prolonged and trustworthy teacher-learner relationships to create a safe and supportive environment. Frameworks such as 'The Educational Alliance' model,⁸¹ the R2C2 model⁵⁴ and The Interpersonal Circumplex⁷⁹ might be helpful in this respect.
- Organise group discussions and ensure shared decision-making. These do not only ease teachers' individual assessment responsibilities but can also improve the assessment outcome.
- Invest in credibility and trustworthiness as quality concepts for stakeholders, the process, and the system. Norcini et al.,³ offer a quality framework for assessment systems.
- Ensure a supportive infrastructure (i.e., available time and resources, effective technology, and sufficient faculty development), while taking the realities of the educational context into account.
- Offer leadership in times of change. Cultural change takes time and, although issues should be addressed quickly, programmatic assessment will not be implemented perfectly from the start.

Concluding remarks

Medical education might still be far from establishing a true learning culture,^{4, 20, 25} however, the implementations of programmatic assessment that were the object of investigation in this dissertation showed promising results. The approach proved instrumental to help move towards a culture that emphasises learning and that supports continuous improvement. However, such a

transition is not without its challenges or hurdles. Although the need for and requirements of the reconceptualisation of competence and the implications for assessment is acknowledged within the literature,⁶² the results showed that, in practice, conventional hegemonies do prevail, most likely in the knowledge domain and in assessment relationships in which teachers' positional agency is exercised dominantly. When the validity of many forms of knowledge and knowing and the limited nature of objectivity and reliability would be embraced in practice, what follows is the need to create conditions in which learners are empowered to position themselves in relation to the assessment regime. In order to utilise the learning potential of programmatic assessment within and beyond formal education, I have argued that the assessment process requires a shift to assessment as a process of co-inquiry to negotiate and determine what counts as legitimate. This should be supported by safe assessment environment and trustworthy assessment relationships in which teachers transform part of their agency towards learners over time and through development. A culture in which critical inquiry is exercised and embraced by all parties involved seems more likely to *ensure* quality of competence and build on valuable knowledge traditions of the field while simultaneously embracing diversity and learner agency to *equip* learners for ongoing learning after formal education. To counterbalance the focus on control and surveillance pushed by current and prevailing conceptualisations of accountability highlighted above, I embrace the call for education to 'push back' the technical managerial approach to accountability^{64, 72} and keep in mind the issue at hand here; to use assessment for meaningful and worthwhile learning and to benefit and support learners long after they leave formal education systems. I believe learners and teacher deserve the trust, support and guidance in reaching their potential. Therefore, I would like to end this dissertation with the words of O'Neill⁷³ who I think captures this spirit poetically, with a call to swing back the pendulum of accountability to an urgent and much needed balance between trust and control.

'Plants do not flourish when we pull them up too often to check how their roots are growing.'

REFERENCES

1. Schuwirth LW, Van der Vleuten CP. Programmatic assessment: From assessment of learning to assessment for learning. *Med Teach*. 2011;33(6):478-85.
2. Taras M. Assessment – Summative and Formative – Some Theoretical Reflections. *Br J Educ Stud*. 2005;53(4):466-78.
3. Norcini J, Anderson MB, Bollela V, Burch V, Costa MJ, Duvivier R, et al. 2018 Consensus framework for good assessment. *Med Teach*. 2018;40(11):1102-9.
4. Watling CJ, Ginsburg S. Assessment, feedback and the alchemy of learning. *Med Educ*. 2019;53(1):76-85.
5. Schuwirth LWT, Van der Vleuten CPM. A history of assessment in medical education. *Adv Health Sci Educ Theory Pract*. 2020;25(5):1045-56.
6. 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):487-98.
7. Bok HG, Jaarsma DA, Spruijt A, Van Beukelen P, Van der Vleuten CP, Teunissen PW. Feedback-giving behaviour in performance evaluations during clinical clerkships. *Med Teach*. 2016;38(1):88-95.
8. Castanelli DJ, Jowsey T, Chen Y, Weller JM. Perceptions of purpose, value, and process of the mini-Clinical Evaluation Exercise in anesthesia training. *Can J Anaesth*. 2016;63(12):1345-56.
9. O'Brien CL, Sanguino SM, Thomas JX, Green MM. Feasibility and Outcomes of Implementing a Portfolio Assessment System Alongside a Traditional Grading System. *Acad Med*. 2016;91(11):1554-60.
10. Li SA, Sherbino J, Chan TM. McMaster Modular Assessment Program (McMAP) Through the Years: Residents' Experience With an Evolving Feedback Culture Over a 3-year Period. *AEM Educ Train*. 2017;1(1):5-14.
11. Zijlstra-Shaw S, Roberts T, Robinson PG. Evaluation of an assessment system for professionalism amongst dental students. *Eur J Dent Educ*. 2017;21(4):e89-e100.
12. 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):654-63.
13. 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):434-44.
14. Griffiths J, Dalgarno N, Schultz K, Han H, Van Melle E. Competency-Based Medical Education implementation: Are we transforming the culture of assessment? *Med Teach*. 2019;41(7):811-8.
15. Bacon R, Kellett J, Dart J, Knight-Agarwal C, Mete R, Ash S, et al. A Consensus Model: Shifting assessment practices in dietetics tertiary education. *Nutr Diet*. 2018;75(4):418-30.
16. Ross S, Binczyk NM, Hamza DM, Schipper S, Humphries P, Nichols D, et al. Association of a Competency-Based Assessment System With Identification of and Support for Medical Residents in Difficulty. *JAMA Netw Open*. 2018;1(7):e184581.
17. Favier RP, Vernooij JCM, Jonker FH, Bok HGJ. Inter-Rater Reliability of Grading Undergraduate Portfolios in Veterinary Medical Education. *J Vet Med Educ*. 2019;46(4):415-22.
18. Imanipour M, Jalili M. Development of a comprehensive clinical performance assessment system for nursing students: A programmatic approach. *Jpn J Nurs Sci*. 2016;13(1):46-54.
19. Pugh D, Regehr G. Taking the sting out of assessment: is there a role for progress testing? *Med Educ*. 2016;50(7):721-9.

CHAPTER 8

20. Harrison C, Wass V. The challenge of changing to an assessment for learning culture. *Med Educ.* 2016;50(7):704-6.
21. Van der Vleuten CP, Schuwirth LW, Driessen EW, Dijkstra J, Tigelaar D, Baartman LK, et al. A model for programmatic assessment fit for purpose. *Med Teach.* 2012;34(3):205-14.
22. 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):276-84.
23. Watling C. The uneasy alliance of assessment and feedback. *Perspect Med Educ.* 2016;5(5):262-4.
24. Altahawi F, Sisk B, Poloskey S, Hicks C, Dannefer EF. Student perspectives on assessment: experience in a competency-based portfolio system. *Med Teach.* 2012;34(3):221-5.
25. Eva KW, Bordage G, Campbell C, Galbraith R, Ginsburg S, Holmboe E, et al. Towards a program of assessment for health professionals: from training into practice. *Adv Health Sci Educ Theory Pract.* 2016;21(4):897-913.
26. 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.
27. Boud D, Molloy E. Rethinking models of feedback for learning: the challenge of design. *Assess Eval High Educ.* 2013;38(6):698-712.
28. Watling C, Ginsburg S, LaDonna K, Lingard L, Field E. Going against the grain: An exploration of agency in medical learning. *Med Educ.* 2021. <https://doi.org/10.1111/medu.14532>
29. 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.
30. Wiliam D. *Embedded formative assessment.* Bloomington, IN: Solution Tree Press; 2011.
31. Chan T, Sherbino J, Mc MAPC. The McMaster Modular Assessment Program (McMAP): A Theoretically Grounded Work-Based Assessment System for an Emergency Medicine Residency Program. *Acad Med.* 2015;90(7):900-5.
32. Watling C, Driessen E, Van der Vleuten CP, Vanstone M, Lingard L. Understanding responses to feedback: the potential and limitations of regulatory focus theory. *Med Educ.* 2012;46(6):593-603.
33. Boud D. Sustainable Assessment: Rethinking assessment for the learning society. *Stud Contin Educ.* 2000;22(2):151-67.
34. Winstone NE, Carless D. *Designing effective feedback processes in higher education: a learning-focused approach.* Abingdon, Oxon: Routledge; 2020.
35. 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):307-20.
36. 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):1492-7.
37. Telio S, Regehr G, Ajjawi R. Feedback and the educational alliance: examining credibility judgements and their consequences. *Med Educ.* 2016;50(9):933-42.
38. Tannock S. No grades in higher education now! Revisiting the place of graded assessment in the reimagination of the public university. *High Educ Stud.* 2015;42(8):1345-57.
39. Deci EL, Koestner R, Ryan RM. Extrinsic Rewards and Intrinsic Motivation in Education: Reconsidered Once Again. *Rev Educ Res.* 2016;71(1):1-27.

40. Laveault D, Allal L. Implementing Assessment for Learning: Theoretical and Practical Issues. In: Laveault D, Allal L, editors. *Assessment for Learning: Meeting the Challenge of Implementation*. Chambridge: Springer; 2016. p. 1-18.
41. Eva KW, Munoz J, Hanson MD, Walsh A, Wakefield J. Which factors, personal or external, most influence students' generation of learning goals? *Acad Med*. 2010;85(10 Suppl):S102-5.
42. Van der Vleuten CPM, Schuwirth LWT, Driessen EW, Govaerts MJB, Heeneman S. Twelve Tips for programmatic assessment. *Med Teach*. 2015;37(7):641-6.
43. Foucault M, Sheridan A. *Discipline and punish: the birth of the prison*. 2nd ed. New York: Vintage Books; 1995.
44. Pearson J. Assessment of agency or assessment for agency? A critical realist action research study into the impact of a processfolio assessment within UK HE preparatory courses for international students. *Educ Action Res*. 2020;29(2):259-75.
45. Leach L, Neutze G, Zepke N. Assessment and Empowerment: Some critical questions. *Assess Eval High Educ*. 2010;26(4):293-305.
46. Boud D, Falchikov N. *Rethinking assessment in higher education: learning for the longer term*. London: Routledge; 2007.
47. Laveault D, Allal L. *Assessment for Learning: Meeting the Challenge of Implementation*. Chambridge: Springer; 2016.
48. Gipps CV. *Beyond testing: towards a theory of educational assessment*. London: The Falmer Press; 1994.
49. Shepard LA. The Role of Assessment in a Learning Culture. *Educ Res*. 2016;29(7):4-14.
50. Pratto F. On power and empowerment. *Br J Soc Psychol*. 2016;55(1):1-20.
51. Van Tartwijk J, Driessen EW. Portfolios for assessment and learning: AMEE Guide no. 45. *Med Teach*. 2009;31(9):790-801.
52. Bates J, Konkin J, Suddards C, Dobson S, Pratt D. Student perceptions of assessment and feedback in longitudinal integrated clerkships. *Med Educ*. 2013;47(4):362-74.
53. Ramani S, Konings KD, Ginsburg S, Van der Vleuten CPM. Twelve tips to promote a feedback culture with a growth mind-set: Swinging the feedback pendulum from recipes to relationships. *Med Teach*. 2019;41(6):625-31.
54. Sargeant J, Lockyer J, Mann K, Holmboe E, Silver I, Armson H, et al. Facilitated Reflective Performance Feedback: Developing an Evidence- and Theory-Based Model That Builds Relationship, Explores Reactions and Content, and Coaches for Performance Change (R2C2). *Acad Med*. 2015;90(12):1698-706.
55. Jamieson J, Hay M, Gibson S, Palermo C. Implementing programmatic assessment transforms supervisor attitudes: An explanatory sequential mixed methods study. *Med Teach*. 2021:1-9.
56. Reynolds M, Trehan K. Assessment: A critical perspective. *High Educ Stud*. 2000;25(3):267-78.
57. Heron J. Assessment revisited. In: Boud D, editor. *Developing Student Autonomy in Learning*. Second ed. London: Kogan Page Limited; 1981. p. 55-68.
58. Pearce J, Tavares W. A philosophical history of programmatic assessment: tracing shifting configurations. *Adv Health Sci Educ Theory Pract*. 2021. DOI: 10.1007/s10459-021-10050-1
59. Hodges BD, Martimianakis MA, McNaughton N, Whitehead C. Medical education... meet Michel Foucault. *Med Educ*. 2014;48(6):563-71.
60. Mann SJ. Alternative Perspectives on the Student Experience: Alienation and engagement. *High Educ Stud*. 2010;26(1):7-19.
61. Kennedy TJ, Regehr G, Baker GR, Lingard LA. 'It's a cultural expectation...!' The pressure on medical trainees to work independently in clinical practice. *Med Educ*. 2009;43(7):645-53.

CHAPTER 8

62. Gipps CV. Socio-cultural aspects of assessment. *Rev Educ Res.* 1999;24:355-92.
63. Kahu ER. Framing student engagement in higher education. *High Educ Stud.* 2013;38(5):758-73.
64. Biesta G. *The beautiful risk of education.* London: Taylor & Francis; 2016.
65. Ecclestone K. Empowerng or Ensnaring?: The Implications of Outcome-based Assessment in Higher Education. *High Educ Q* 2002;53(1):29-48.
66. Taras M. Issues of power and equity in two models of self-assessment. *Teach High Educ.* 2008;13(1):81-92.
67. Cruess RL, Cruess SR. Expectations and obligations: professionalism and medicine's social contract with society. *Perspect Biol Med.* 2008;51(4):579-98.
68. Govaerts MJB, Van der Vleuten CPM, Holmboe ES. Managing tensions in assessment: moving beyond either-or thinking. *Med Educ.* 2019;53(1):64-75.
69. Ecclestone K, Pryor J. 'Learning Careers' or 'Assessment Careers'? The Impact of Assessment Systems on Learning. *Br Educ Res J.* 2003;29(4):471-88.
70. Ajjawi R, Molloy E, Bearman M, Rees CE. Contextual Influences on Feedback Practices: An Ecological Perspective. In: Carless D, Bridges SM, Chan CKY, Glofcheski R, editors. *Scaling up Assessment for Learning in Higher Education.* Vol. 5. Singapore: Springer; 2017. p.129-43.
71. Broadfoot P. *The sociology of assessment: comparative and policy perspectives: the selected works of Patricia Broadfoot.* Abingdon, Oxon: Routledge; 2021.
72. Biesta G. Good education in an age of measurement: on the need to reconnect with the question of purpose in education. *Educ Assess Eva Acc.* 2008;21(1):33-46.
73. O'Neill O. *A question of trust.* Cambridge: Cambridge University Press; 2002.
74. Van der Vleuten CP. The assessment of professional competence: Developments, research and practical implications. *Adv Health Sci Educ Theory Pract.* 1996;1(1):41-67.
75. Govaerts M, Van der Vleuten CP. Validity in work-based assessment: expanding our horizons. *Med Educ.* 2013;47(12):1164-74.
76. Hodges B. Medical education and the maintenance of incompetence. *Med Teach.* 2006;28(8):690-6.
77. Airasian PW. Measurement Driven Instruction: A Closer Look. *Educ Meas.* 1988;7(4):6-11.
78. Stiggins R. Two Disciplines of Educational Assessment. *Meas Eval Couns Dev.* 1993;26(1):93-104.
79. Gurtman MB. Exploring Personality with the Interpersonal Circumplex. *Soc Psychol Compass.* 2009;3(4):601-19.
80. Van Tartwijk J, Veldman I, Verloop N. Classroom management in a Dutch teacher education program: a realistic approach. *Teach Educ.* 2011;22(2):169-84.
81. Telio S, Ajjawi R, Regehr G. The "educational alliance" as a framework for reconceptualizing feedback in medical education. *Acad Med.* 2015;90(5):609-14.
82. Govaerts MJ, Van der Vleuten CP, Schuwirth LW, Muijtjens AM. Broadening perspectives on clinical performance assessment: rethinking the nature of in-training assessment. *Adv Health Sci Educ Theory Pract.* 2007;12(2):239-60.



ADDENDUM

Summary

Assessment is often a burden to those involved in education. It is believed to have a strong impact on learning, however, neither necessarily desirable nor positive. Assessment can lead to surface learning approaches, strategic choices, teaching and learning to the test, anxiety, stress, and fear of failing. The ambitions to overcome the undesirable and unintended consequence of assessment and to use assessment for more than accountability are high. Specifically, the assessment challenges accompanied with competency-based medical education (CBME) have led to the conceived consequential need to simultaneously use assessment of learning with assessment for learning. In this dissertation, three empirical studies, a knowledge synthesis, and two commentary papers, are included, which have sought to further explore and better understand the mechanisms explaining the complex relationship between assessment and learning within CBME. Overall, the dissertation sets out to gain an in-depth understanding of a whole-system assessment approach coined as programmatic assessment and the concept of assessment as a continuum of increasing stakes within such a system from a sociocultural perspective and an interpersonal perspective.

The first two empirical studies explored the meaning and experiences with assessment stakes and programmatic assessment from individual perspectives. **Chapter 2** explored how learners perceive assessment stakes within programmatic assessment and which factors influence these perceptions. Data were gathered by interviewing twenty-six learners from three different countries and five different programmes, ranging from undergraduate to postgraduate medical education. The interviews explored learners' experience with and perception of assessment stakes. An open and qualitative approach to data gathering and analyses inspired by the constructivist grounded theory approach was used to analyse the data and reveal underlying mechanisms influencing learners' perceptions. Learner agency emerged from the analysis as key for understanding learners' perception of assessment stakes. Several design factors of the assessment programme afforded learners' opportunities to exercise agency over the assessment experience, mainly the opportunities to influence assessment outcomes, to collect evidence and to improve. Teachers played a powerful role in learners' assessment perceptions. In this study, factors were identified that influenced learners' perception of assessment stakes and which can help the design of assessment programmes in which assessment supports learning and competency development. As teachers appeared to play a key role in learners' assessment perceptions, the study in **Chapter 3** turned to the perspective of teachers. The study explored how teachers conceptualise assessments within programmatic assessment and how they engaged with learners in assessment relationships. Twenty-three interviews were conducted at two different graduate-entry medical training programmes following a theoretical sampling approach. Results showed that teachers conceptualised the meaning and purpose of 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 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 while 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. The study concluded with the notion that teachers believed 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 hold significant stakes for teachers themselves. Important design features were identified that support teachers and preserve the benefits of prolonged engagement in assessment relationships, which were the sampling across different assessments and assessors, and the introduction of progress committees.

The third empirical study presented in **Chapter 4** took an interpersonal perspective and investigated assessment perceptions through the lens of teacher-learner relationships. A focussed analysis using sensitising concepts from interpersonal theory was conducted to elucidate the influence of the teacher-learner relationship on learners' assessment perceptions in five different settings of programmatic assessment. The study showed a strong relation between learners' perceptions of the teacher-learner relationship and their assessment task perception. Two important sources for the perception of teachers' agency emerged from the data: positional agency and expert agency. Together with teacher's communion level, both types of teachers' agency are important for understanding learners' assessment perceptions. High levels of teacher communion had a positive impact on the perception of assessment for learning, in particular in relations in which teachers' agency was less dominantly exercised. When teachers exercised these sources of agency dominantly, learners felt inferior to their teachers, which could hinder the learning opportunity. This work showed how agency is negotiated within interpersonal assessment relationships and when and how this is done effectively in order to stimulate assessment for learning. The Interpersonal Circumplex offers a useful reference for understanding assessment relationships and provides opportunities for faculty development that help teachers develop positive and productive relationships with learners in which the potential of low-stakes assessments for learning is utilised.

Chapter 5 and **Chapter 6** offer two commentary papers. The first paper reflects on the role of and consequences for remediation in educational pathways when taking on a programmatic approach to assessment. The second paper offers a perspective on the decision-making processes of assessment in a whole-system approach and questions the role of learner agency in this process. In this chapter the way in which heuristics and other decision-making processes are commonly considered in the literature was juxtaposed. The paper calls to reconsider the role of learners in this high-stakes process, especially if one aims to benefit self-regulated learning.

The understanding gained from the empirical work guided the systematic integrated review presented in **Chapter 7**. Conducting a knowledge synthesis allowed for further finetuning as insight was gained in the underlying mechanisms based on other implementations and researchers' interpretations of their impact in practice. Characteristics of the twenty-seven included studies, which used a variety of methods and data sources, were extracted and

ADDENDUM

synthesised using descriptive statistics and thematic analysis. Programmatic assessment was found to enable meaningful triangulation for robust decision-making and used as a catalyst for learning. However, several problems were identified, including overload in assessment information and the associated workload, counterproductive impact of using strict requirements and summative signals, lack of a shared understanding of the nature and purpose of programmatic assessment, and lack of supportive interpersonal relationships. Thematic analysis revealed that the success and challenges of programmatic assessment were best understood by the interplay between quantity and quality of assessment information, and the influence of social and personal aspects on assessment perceptions. The knowledge synthesis led to the conclusion that although some of the evidence was compelling to support the effectiveness of programmatic assessment in practice, tensions will emerge when simultaneously stimulating the development of competencies and assessing its result. The paper offers inferred strategies from the literature that can provide guidance for navigating these tensions.

Finally, in the general discussion presented in **Chapter 8**, the results of the individual studies were summarised and synthesised to answer the overall research question and to contribute to the academic debate on the use of assessment for learning in a context in which current conceptualisations of accountability risk dominating the assessment culture. The main conclusion of this dissertation was that assessment for learning in a high-stakes context requires careful consideration of three concepts: agency, trustworthy assessment relationships, and critical inquiry. It was argued that the importance of learner agency in assessment is not only a logical consequence of the social-interpretivist paradigm but based on the research findings presented in this dissertation, is key to enable the dual-purposes aspired with the introduction of programmatic assessment. Assessment that potentially benefits learners beyond formal education calls for the democratisation of the assessment process, which implicates a shift to assessment as a process of co-inquiry to negotiate and determine what counts as legitimate. A culture in which critical inquiry is exercised and embraced by all parties involved is more likely to ensure quality of competence and build on valuable knowledge traditions of the field while simultaneously embracing diversity and learner agency to equip learners for ongoing learning after formal education.

Impact Paragraph

The impact paragraph offers a reflection on anticipated and achieved social and scientific impact of this dissertation and the target groups that are thought to benefit from the results. Furthermore, it offers an overview of the dissemination activities undertaken throughout the duration of the research project. I do however believe that achieved impact is a difficult claim to make, the following will refer more to the intended impact than the ‘actual’ achieved impact. If anything, conducting this research first and foremost has had a major impact on my own personal and professional development. For the rest, the proof is in the pudding and, in line with the spirit of this dissertation, the perceptions of the target groups matter.

Main objectives and the most important results and conclusions (Research)

Assessment is believed to have a strong impact on learning. However, risks associated with negative and undesirable consequences of assessment are significant and can even be dramatic. The assessment challenges accompanied with competency-based medical education (CBME) have led to the conceived consequential need to simultaneously use assessment of learning with assessment for learning. This dissertation aimed to gain an in-depth understanding of a whole-system assessment approach from a sociocultural and an interpersonal perspective. It gives accounts of the ways in which learners and teachers in CBME recognised, processed, and utilised assessment for learning within a whole-system approach of assessment, and synthesises the findings of other scholars investigating such an approach in practice. Three empirical studies, a knowledge synthesis, and two commentary papers were included, which have sought to explain the complex relationship between assessment and learning. Programmatic assessment was found to enable meaningful triangulation for robust decision-making and used as a catalyst for learning. However, several problems were identified, including overload in assessment information and the associated workload, counterproductive impact of using strict requirements and summative signals, lack of a shared understanding of the nature and purpose of programmatic assessment, and lack of supportive interpersonal relationships. With this dissertation the affordances of a whole-system approach were further clarified, which concerned the use of multiple assessments in a coherent manner, characteristics of assessment information for learning and the role of follow-up activities, prolonged teacher-learner relationships, adaptive guidance and support, and the introduction of an independent assessment committee. The main conclusion of this dissertation was that assessment for learning in a high-stakes context requires careful consideration of three concepts: agency, trustworthy assessment relationships, and critical inquiry. In order to utilise the learning potential of programmatic assessment, it was argued that the assessment process requires a shift to assessment as a process of co-inquiry to negotiate and determine what counts as legitimate. This should be supported by a safe assessment environment and trustworthy interpersonal relationships in which teachers transform part of their agency towards learners over time and through development. An assessment culture in which critical inquiry is exercised and embraced by all parties involved is more likely to *ensure* quality of competence and build on valuable knowledge traditions of the field, while simultaneously embracing learner agency and diversity to *equip* learners for ongoing learning after formal education.

The (potential) contribution of the results to science, social sectors and social challenges (Relevance & Target group)

The results of this dissertation are relevant on three levels, which involve multiple stakeholders: the (inter)national level concerning policy makers and society-at-large, the institutional level concerning programme directors and curriculum designers, and the programme level concerning staff and students. Assessment influences all those involved in education. Educational institutes allocate, often scarce, educational resources to the assessment process, such as staff time and technical and managerial support. Societal expectations towards accountability may never have been more pressing, and ambitions to overcome undesirable consequences of assessment and to use assessment for more than accountability are high. Despite this high interest and relevance of the topic, research from a sociocultural perspective is still scarce, which is problematic given the shifting conceptualisations concerning learning and assessment in medical education. The results from this dissertation contribute to the academic debate on the use of assessment for learning in a context in which current conceptualisations of accountability risk dominating the assessment culture and where discourses concerning assessment and learning are changing. By using a sociocultural and interpersonal perspective on the complex relationship between assessment and learning, the dissertation contributes to this shift in discourse concerning assessment in competency-based medical education. The gained insights concerning the affordances of programmatic assessment and the conditions that led to these emergences, serve to illuminate, and provide a meaningful basis for moving the theory and practice of a whole-system assessment approach forward.

The aspirations to use assessment for more than accountability is growing in countries across Europe and beyond. Currently, our local institute, Maastricht University, is updating its vision on assessment and explicitly aims for such aspirations, calling its new policy document 'Moving from an assessment culture of testing toward a culture of feedback and development'. Around the world, educational institutes are rapidly and increasingly implementing whole-system approaches to assessment. Overcoming undesirable and unintended consequences of assessment is of interest and at stake for all those involved in education and assessment, specifically due to the negative impact which can marginalise students and have a long-lasting and dramatic impact on their life's, such as burnout problems and anxieties. This dissertation aims to provide more insight in the mechanisms underlying the complex relationship between assessment and learning and to benefit stakeholders with these insights. Specific attention was given to successful implementation characteristics, perceptions of key stakeholders, and ways in which higher education institutions and teaching hospitals can promote an assessment culture supporting learning and longitudinal competency development, which is characterised by a safe and supportive environment, prolonged trustworthy relationships, and promotes ownership and learning agency. Such a culture is assumed to benefit more positive and desirable effects than summative assessment cultures, which overemphasise standardisation, lack of trust, and technical managerial control measures. This dissertation offers strategies from research and existing literature to provide support in navigating tensions when assessment is aspired to fulfil a double duty. Furthermore, implications and suggestions that appertain to faculty development are provided.

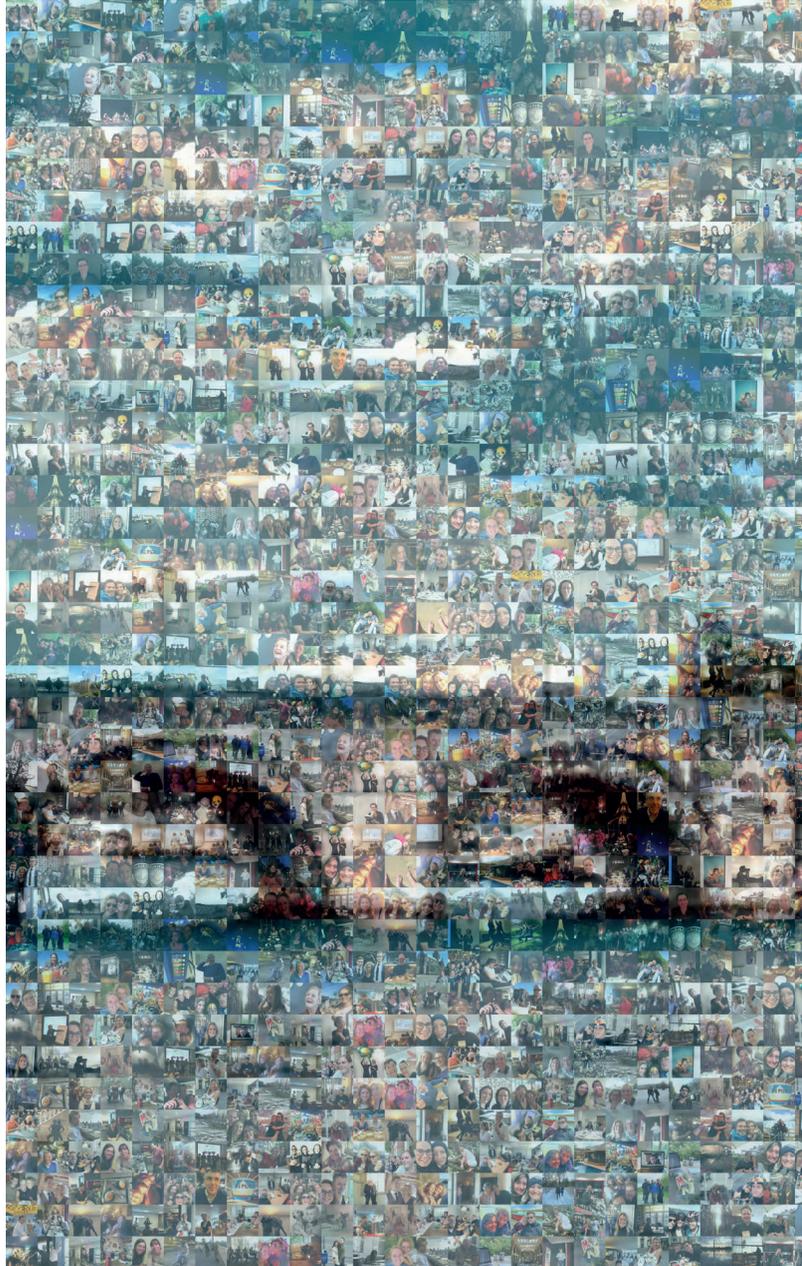
Dissemination (Activity)

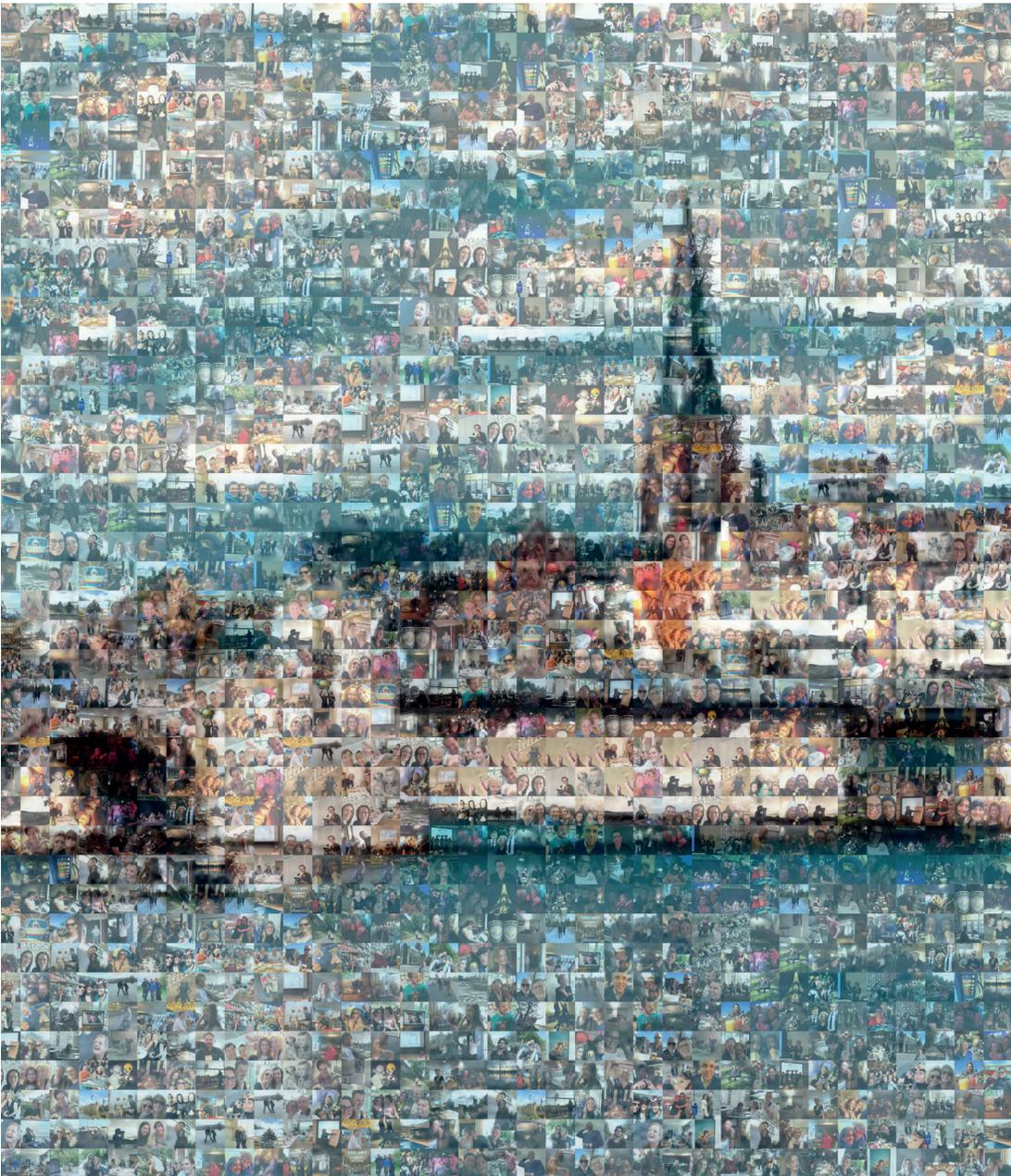
All studies included in this dissertation are available as open access published manuscripts in scientific journals that address a broad audience in the field of medical and health sciences education and assessment specifically. The dissemination of these studies was further enhanced by research paper presentations at national and international scientific conferences for medical and health science education (i.e., The Association for Medical Education in Europe (AMEE), The European Board of Medical Assessors (EBMA), Assessment of Competence in Medical and the Health Professions (OTTAWA), and The Dutch Association for Medical Education (NVMO)). Moreover, the knowledge synthesis was highlighted in an invited keynote lecture I delivered for The European Board of Medical Assessors in 2019 and served as input for the symposium I have organised at the biannual conference for Assessment of Competence in Medical and the Health Professions (OTTAWA) in 2020. Furthermore, conducting this dissertation influenced my work and activities as an educational consultant in providing faculty development concerning assessment at the Faculty of Health, Medicine and Life Sciences at Maastricht University and at the Maastricht University Medical Centre, and as a teacher and coordinator in the master programme of the School of Health Professions Education.

Since the publication of the papers, I have been invited to present and discuss the results at locally organised educational activities (e.g., the journal club of General Practitioners at the Leiden University Medical Centre, the Wilson Centre of the University of Toronto, Canada, the Medical Education Grand Rounds at the University of Virginia School of Medicine) and have been involved in consultancy projects concerning the design and implementation of programmatic assessment or in support of the preparation for accreditation committees (e.g., Sydney Medical School, Australia, and the University of Fribourg, Switzerland). I have developed and delivered workshops on the above-mentioned conferences (NVMO, AMEE) and a two-day workshop programme funded by the National Board of Medical Examiners of the United States of America (NBME) for a national Faculty Development project for seven different medical school in Brazil. Overall, these invitations highlight the visibility of and interest in the findings of this dissertation.

Acknowledgements

To all the individual 'snapshots' of wonderful moments and amazing people who have coloured and shaped my 'whole' experience. None of this would have been possible without you. Joie de Vivre!





About the Author

Suzanne Schut was born in The Hague, the Netherlands on October 20th, 1982. After graduating with honours from the Educational Science programme in Utrecht in 2010, she started working as a teacher and consultant in higher education. In 2014, she joined the medical education community, starting at the Leiden University Medical Centre (LUMC) and then at the Faculty of Health, Medicine and Life Sciences of Maastricht University. She got the opportunity to pursue her research interest at the School of Health Professions Education (SHE) where she was able to combine her job as assessment consultant with teaching and writing a PhD dissertation. Her main activities in the Department of Educational Development and Research are focussed on student assessment. She is appointed as educational advisor within the Assessment Taskforce and responsible for the knowledge assessment domain. She is chair of the assessment review committee and has been involved in educational development and design, teaching and faculty development. Her research interest are competency-based education and assessment, and more specifically student learning and teacher-learner relationships within the assessment environment.

SHE Dissertation Series

The SHE Dissertation Series publishes dissertations of PhD candidates from the School of Health Professions Education (SHE) who defended their PhD theses at Maastricht University. The most recent ones are listed below. For more information go to: <https://she.mumc.maastrichtuniversity.nl>

- Schillings, M. (06-07-2021) Talking about feedback: Face-to-face peer dialogue about written feedback
- Wilbur, K. (05-07-2021) NO WHERE | NOW HERE: Context and Competency Expectations in Workplace-Based Training
- Bendermacher, G. (02-07-2021) Navigating from Quality Management to Quality Culture
- Ahmed Khan, R. (29-06-2021) Assessing curriculum viability in Undergraduate Medical Education
- Chim, H.Q. (30-03-2021) Physical Activity Behavior and Learning in Higher Education
- Dominguez, L.C. (23-02-2021) Persistence in surgical training: The role of job crafting and leadership
- Bindels, E. (22-02-2021) DOING WELL, GETTING BETTER; Facilitating physicians' reflection on their professional performance
- Iqbal, Z. (15-12-2020) All stakeholders matter in faculty development: Designing entrustable professional activities for small group facilitation
- Tran, QT. (09-12-2020) Nationwide implementation of medical skills training laboratories in a developing country: studies from Vietnam
- Pacifico, J. (30-11-2020) Making the Implicit Explicit: Uncovering the Role of the Conceptions of Teaching and Learning and the Perceptions of the Learning Climate in Postgraduate Medical Training
- Nishigori, H. (17-11-2020) Why do doctors work for patients? Medical professionalism in the era of neoliberalism
- Oudkerk Pool, A. (06-11-2020) Competency-based portfolio assessment – Unraveling stakeholder perspectives and assessment practices
- Geel van, K. (05-11-2020) Lifelong learning in radiology: all eyes on visual expertise
- Stammen, L. (16-10-2020) Pursuing - High-Value, Cost-Conscious Care - The Role of Medical Education
- Meulen van der, M. (15-10-2020) Assessment of physicians' professional performance using questionnaire-based tools
- Matsuyama, Y. (05-10-2020) Contextual attributes fostering self-regulated learning in a teacher-centered culture: learner's professional identity formation is a trigger
- Rovers, S. (16-09-2020) Growing knowledge, supporting students' self-regulation in problem-based learning
- Bourgeois-Law, G. (03-09-2020) Conceptualizations of remediation for practicing physicians
- Giuliani, M. (19-05-2020) A Critical Review of Global Curriculum Development, Content and Implementation in Oncology
- Schreurs, S. (20-03-2020) Selection for medical school; the quest for validity
- Schumacher, D. (19-03-2020) Resident Sensitive Quality Measures: Defining the Future of Patient-Focused Assessment
- Schlbach, C. (21-02-2020) To be continued.... Supporting physicians' lifelong learning
- Kikukawa, M. (17-12-2019) The situated nature of validity: Exploring the cultural dependency of evaluating clinical teachers in Japan
- Kelly, M. (10-12-2019) Body of knowledge. An interpretive inquiry into touch in medical education
- Klein, D. (06-11-2019) The performance of medical record review as an instrument for measuring and improving patient safety

‘Professor? Is this all real?
Or is it just happening in my head?’

‘Of course it’s happening inside your head, Harry.
Why should that mean it is not real?’

Harry Potter and the Deathly Hollows



