



OPEN ACCESS

EDITED BY
Mustafa Asil,
University of Otago, New Zealand

REVIEWED BY
Dario Torre,
University of Central Florida,
United States
Chris Roberts,
The University of Sydney, Australia

*CORRESPONDENCE
Liesbeth Baartman
liesbeth.baartman@hu.nl

SPECIALTY SECTION
This article was submitted to
Assessment, Testing and Applied
Measurement,
a section of the journal
Frontiers in Education

RECEIVED 29 April 2022
ACCEPTED 13 September 2022
PUBLISHED 04 October 2022

CITATION
Baartman L, van Schilt-Mol T and van
der Vleuten C (2022) Programmatic
assessment design choices in nine
programs in higher education.
Front. Educ. 7:931980.
doi: 10.3389/feduc.2022.931980

COPYRIGHT
© 2022 Baartman, van Schilt-Mol and
van der Vleuten. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

Programmatic assessment design choices in nine programs in higher education

Liesbeth Baartman ^{1*}, Tamara van Schilt-Mol ² and
Cees van der Vleuten ³

¹Research Group Vocational Education, Research Centre for Learning and Innovation, Utrecht University of Applied Sciences, Utrecht, Netherlands, ²School of Education, HAN University of Applied Sciences, Nijmegen, Netherlands, ³School of Health Professions Education, Maastricht University, Maastricht, Netherlands

In programmatic assessment (PA), an arrangement of different assessment methods is deliberately designed across the entire curriculum, combined and planned to support both robust decision-making and student learning. In health sciences education, evidence about the merits and pitfalls of PA is emerging. Although there is consensus about the theoretical principles of PA, programs make diverse design choices based on these principles to implement PA in practice, fitting their own contexts. We therefore need a better understanding of how the PA principles are implemented across contexts—within and beyond health sciences education. In this study, interviews were conducted with teachers/curriculum designers representing nine different programs in diverse professional domains. Research questions focused on: (1) design choices made, (2) whether these design choices adhere to PA principles, (3) student and teacher experiences in practice, and (4) context-specific differences between the programs. A wide range of design choices were reported, largely adhering to PA principles but differing across cases due to contextual alignment. Design choices reported by almost all programs include a backbone of learning outcomes, data-points connected to this backbone in a longitudinal design allowing uptake of feedback, intermediate reflective meetings, and decision-making based on a multitude of data-points made by a committee and involving multi-stage procedures. Contextual design choices were made aligning the design to the professional domain and practical feasibility. Further research is needed in particular with regard to intermediate-stakes decisions.

KEYWORDS

programmatic assessment, curriculum design, feedback, formative, summative

Introduction

In higher (professional) education, students often experience a “testing culture,” involving a high number of summative assessments and a culture of teaching and learning to the test (Frederiksen, 1984; Jessop and Tomas, 2017). In reaction to this testing culture, insights from practice, theory and research guided the design of a new approach to assessment, called Programmatic Assessment (van der Vleuten and Schuwirt, 2005; van der Vleuten et al., 2012; Torre et al., 2020). Programmatic Assessment (PA) entails a fundamental paradigm shift in our approach to assessment, both from a learning perspective and a decision-making perspective. PA involves the longitudinal collection of so-called “data-points” about student learning, for example by assignments, feedback from peers, supervisors or clients, and observations in practice. In PA, the arrangement of different data-points is deliberately designed across the entire curriculum, combined and planned to optimize both robust decision making and student learning (van der Vleuten et al., 2012; Heeneman et al., 2021).

A growing body of evidence to support the value of programmatic assessment is emerging (Schut et al., 2020a). In the domain of health sciences education, many curricula and corresponding assessment programs worldwide have been redesigned according to the principles of PA (e.g., Wilkinson et al., 2011; Bok et al., 2013; Jamieson et al., 2017). In this domain, research has shown that PA can generate sufficient information to enable robust decision-making and can be a catalyst for learning (e.g., Driessen et al., 2012; Heeneman et al., 2015; Imanipour and Jalili, 2016; Schut et al., 2020a).

However, more empirical research is needed focusing on the implementations and effects of PA in wider educational practices. In most educational programs, using assessment information to steer learning in a continuous flow of information and using this information for robust decision-making is not self-evident nor common practice. This can partly be explained by the design choices made. Often curricula are divided into separate modules or subjects, leading to misalignment between curricular objectives and assessment. Consequently, assessments concern module objectives instead of curricular objectives (Kickert et al., 2021). Within each subject, fragmentation continues by deconstructing objectives into smaller assessable elements, thereby further losing track of the whole (Sadler, 2007), leading to a lack of understanding of the coherence of the curriculum by students. As a consequence, feedback and learning experiences from one module are not used by students for the next, or feedback is even ignored (Harrison et al., 2016).

Since intentional design choices can thus determine to a large extent whether learning processes are stimulated (van den Akker, 2004; Carvalho and Goodyear, 2018; Bouw et al., 2021) and whether robust decision making is possible, PA is often described as a design issue (Torre et al., 2021).

Although there is consensus about the theoretical principles that serve as a basis for the design of PA, programs make various specific design choices based on these principles fitting with their own context. After all, curriculum design choices are highly context-specific (Heeneman et al., 2021). We therefore need a better understanding of design choices made by different programs, the considerations about these design choices and the effects on both teachers and students. This involves an exploration of variability of design choices made by different programs regarding the principles of PA, considering their local context. In this study, we therefore interviewed teachers/curriculum designers of nine programs in different professional domains (Table 1): teaching, health sciences, paramedics, and communication. In all cases, PA has been implemented. The research questions of this study were: (1) what design choices are made by the different programs, (2) to what extent do these design choices fit the principles of PA, (3) what are student and teacher experiences, that is, how does the design work out in practice, and (4) what differences can be found between the nine participating programs?

Theoretical framework: Principles of programmatic assessment

Programmatic Assessment is founded on a number of theoretical frameworks about validity and reliability (van der Vleuten et al., 2010) and learning theories (Torre et al., 2020). Implementations of PA vary—as we will show in this study—but some universal principles of PA have been distilled and agreed upon (Heeneman et al., 2021). In their consensus paper about the principles of PA, Heeneman et al. (2021) present 12 agreed-upon principles (merged into 8 groups of principles). In a complementary paper about manifestations about these principles in practice, Torre et al. (2021) present three thematic clusters covering various principles: (1) continuous meaningful feedback to promote dialogue, (2) mixed methods across and within a continuum of stakes, and (3) establishing equitable and credible decision-making processes. Given that the different principles of PA have a high degree of interrelatedness, for this study we decided to use six principles as the starting point for data-analysis. These six principles were formulated based on earlier work on PA (e.g., van der Vleuten et al., 2012; Jamieson et al., 2017; de Jong et al., 2019; Schut et al., 2020a), as the consensus statement was published after data collection and data analysis of the current study. The six principles used in this study, however, cover all 12 principles of the consensus statement (Heeneman et al., 2021), as summarized in Table 1. Some principles were grouped together, like principles 10, 11, and 12 (recurrent meetings with mentors involving review, dialogue, and working toward self-regulation), and principles 6, 7, and 8 (on decision-making, connected to a framework of learning outcomes). In our study, we chose to explicitly

distinguish a principle regarding a framework or backbone of learning outcomes (PA principle 1), as our practical experiences taught us this backbone is an important starting point for further design choices. In the next paragraphs, the six principles of PA used in this study are shortly described.

Programmatic assessment principle 1: A framework of learning outcomes makes up the backbone of the program, providing the basis for data-points and aggregation

In PA, all data-points are connected to a “backbone” or a framework of learning outcomes, for example a set of competences or complex skills specified in a number of levels to be attained at the end of each phase of the educational program (e.g., a year). The choice of data-points is guided by constructive

TABLE 1 Comparison of the PA principles used in the current study and the consensus statement by Heeneman et al. (2021).

	PA principle used in this study	Related principles in Heeneman et al. (2021)
1	A framework of learning outcomes makes up the backbone of the program, providing the basis for data-points and aggregation	5, 8, 10
2	A mix of different data-points is used to gain insight in student development	1, 4, 5
3	All data-points focus on optimizing feedback and do not involve pass/fail decisions	2, 3
4	A continuous dialogue about data-points, feedback, and self-regulation	10, 11, 12
5	There is a proportional relationship between the number of data-points and the stakes of the decision	6, 7, 8
6	The higher the stakes of the decision, the more assessor expertise is needed	7, 9

PA principles Ottawa consensus paper (Heeneman et al., 2021).

- Every (part of an) assessment is but a data-point.
- Every data-point is optimized for learning by giving meaningful feedback.
- Pass/fail decisions are not given on a single data-point.
- There is a mix of methods of assessment.
- The method chosen depends on the educational justification.
- The distinction between summative-formative is replaced by a continuum of stakes.
- Decision-making on learner progress is proportionately related to the stake.
- Assessment information is triangulated across data-points toward an appropriate framework.
- High-stakes decision are made in a credible and transparent manner, using a holistic approach.
- Intermediate review is made to discuss and decide with the learner on their progress.
- Learners have recurrent meetings with mentors/coaches using self-analysis of all assessment data.
- PA seeks to gradually increase the learner's agency and accountability for their own learning.

alignment (Biggs, 1996), that is methods of assessment should be in line with intended learning objectives and curriculum activities. This implies that data-points should be in line with this backbone of learning outcomes (Heeneman et al., 2021). Data-points are thus learning and assessment activities and materials that inform the backbone, and decision-making is based on aggregating information from the data-points toward this backbone.

Programmatic assessment principle 2: A mix of different data-points is used to gain insight in student development

Any assessment method has its limitations in terms of validity, reliability, and impact on learning (van der Vleuten, 1996). In PA, therefore, a mix of different data-points is used to gain insight in student development and enable continuous monitoring. Data-points can involve any performance relevant information. It can be assessment information (written and oral testing, self-peer assessment, direct observation of simulated, or real performance) or any artefact of an activity (log book, video, infographic). Related to principle 1, data-points are chosen based on their fit to the backbone. Assessment activities drive learning, so besides the purpose of decision-making based on data-points, the educational justification for the choice of data-points is really important. Here, a mix of data-points means that students can practice and show progress in diverse manners, sometimes in writing, other times in practical assignments, interaction with others, etc.

Programmatic assessment principle 3: All data-points focus on optimizing feedback and do not involve pass/fail decisions

In PA, single data-points are viewed as not fit for making pass/fail decisions. The purpose of individual data-points is to provide meaningful feedback to the student. This focus on feedback was recognized as one of the most important components of PA in the consensus statement (Heeneman et al., 2021). It allows students to learn from their mistakes and use feedback for subsequent tasks, without fearing the consequences of not “passing” a test. PA aims at continuous feedback, as data-points provide a longitudinal flow of information about student progress. The feedback loop is closed as students have to take feedback on board by reflecting on it and demonstrate improvement in subsequent data-points (Torre et al., 2020). For giving feedback on complex skills (communication, professionalism, collaboration) narrative feedback is preferred (Ginsburg et al., 2017).

Programmatic assessment principle 4: A continuous dialogue about data-points, feedback, and self-regulation

Data-points are continuously collected throughout the learning process, analyzed by the student and mentor, enabling self-regulated learning, and learner ownership (Heeneman et al., 2015; Schut et al., 2020b). Students are supported to interpret assessment information and act upon it in subsequent tasks (Boud and Molloy, 2013; Baartman and Prins, 2018). PA seeks to gradually increase students' agency for their own learning process and progress, starting from rather structured and curriculum/teacher-controlled assessment to self-owned assessment and learning (Schut et al., 2018; Torre et al., 2020). Also, feedback is most effective when it is a loop, a cyclic process involving dialogue (Carless et al., 2011; Boud and Molloy, 2013; Gulikers et al., 2021). Recurrent meetings with mentors/coaches need to support self-analysis of all assessment data (Heeneman and de Grave, 2017) and students' feedback literacy needs to be supported and gradually developed (Price et al., 2011; Schut et al., 2018).

Fundamental to PA is the idea of meaning making: students analyze and triangulate all information (data-points) available to them so far, identifying strengths and weaknesses. Students actively construct meaning of this information. Here, dialogue between students and teachers in the role of mentor/coach is necessary: learning is seen as a social activity, as shared meaning making between stakeholders (de Vos et al., 2019). Mentors/coaches are regular teachers who engage in the role of coaching the learner. Data-points create various meaningful perspectives. Understanding of these perspectives occurs in a social and collaborative setting. By having these reflections and discussions with a mentor/coach the use of feedback and self-directed learning is promoted. Furthermore, the student builds up a trusted relationship with a mentor/coach, which adds to creating a feedback culture (Ramani et al., 2020; Schut et al., 2020b).

Programmatic assessment principle 5: There is a proportional relationship between the number of data-points and the stakes of the decision

In PA, the student gathers data-points and makes reflections on progress. All this information is stored in a(n) (electronic) portfolio, which is used for decision-making. This 5th principle of PA describes the proportional relationship between the number of data-points in the portfolio and the stakes of the decision, which is conceptualized as a continuum from low to high stakes. Low-stakes means that the decision has limited

consequences for the student, for example feedback on an assignment might entail that the student has to improve the assignment. High-stakes decisions on the other hand have important consequences for students, such as graduation or promotion to a next phase of the program. In between, intermediate-stakes decisions focus on remediation and setting new learning goals together with the student.

Decision-making in PA is proportional. The stakes of an assessment decision are proportional to the richness of information in the portfolio on which it is based (Heeneman et al., 2021). To make high-stakes decisions, information about student learning is collected until saturation of information is reached (de Jong et al., 2019). More time to collect and collate multiple and diverse data-points allows for the inclusion of more different perspectives. In PA, assessment information that pertains to the same content is triangulated, to constructs such as competences or complex skills. As said earlier (principle 1), an appropriate framework is needed to structure triangulation.

Programmatic assessment principle 6: The higher the stakes of the decision, the more assessor expertise is needed

High-stakes decisions about students are based on the interpretation and aggregation of a multitude of data-points and expert group-decision making procedures (de Jong et al., 2019; Torre et al., 2020). High-stakes decisions are made by a committee of experts. Based on the data-points, they judge whether a student has reached the intended outcomes and make a pass/fail decision. For most students the information and the progress is crystal clear, and decision-making by the committee is not difficult. However, a small part of the students may have not or barely reached the learning outcomes. In these cases, the committee deliberates, weighs information, continues the discussion until a consensus is reached on a pass or a fail. One may say that the more it matters, more assessor expertise is used to come to a judgment: the assessor expertise used is proportional to the clarity of information. The decision of the committee should not come as a big surprise to the student. When it does, something has gone wrong in the previous feedback cycles. Decision making is based on the expertise of multiple assessors who are able to justify their judgment and, moreover, thoroughly substantiate their judgments (Oudkerk Pool et al., 2018).

These theoretical principles of PA can be realized in many different manifestations in practice. Different design choices can be made with regard to these six principles, for example regarding the way guidance/mentoring is organized, which people are involved in making high-stakes decisions and how feedback giving and seeking is organized. PA-systems need to be made fit-for-purpose given the local contextual factors. In this study, we therefore interviewed representatives of nine

programs from different professional domains to explore the design choices made, their adherence with the six principles of PA, reported student and teacher experiences on how the design choices work out in practice, and potential differences between the programs.

Materials and methods

Participants

In this exploratory multiple case study, participants were 13 teachers/curriculum designers of nine different higher education programs, including bachelor, master, and teacher professionalization at the workplace. In two interviews, team leaders and/or students participated as well. Some teachers/curriculum designers were interviewed twice (for an overview, see [Table 2](#)). The participants can be seen as information-rich informants regarding their program. Participants and programs were selected through the network of the authors. At the time of data collection, not many programs in the Netherlands had implemented PA and gained practical experiences with PA. Key persons in our network were contacted and first short interviews allowed us to judge whether the program actually (seemed to) fit the principles of PA and whether the participant(s) could provide rich information about the program. In this stage, one program was discarded from further data collection and whenever necessary additional participants were asked to participate in the interviews. The unit of analysis (the case) is each program implementing PA ([Yazan, 2015](#); [Cleland et al., 2021](#)). The participants are seen as representatives of their particular program (though voicing their own opinion) and all had in-depth knowledge and experience

with regard to design and implementation. In seven cases, PA had been implemented for several years, in the other two cases the designed curriculum was about to be implemented at the time of the interview.

Data collection

Open (group) semi-structured interviews were conducted, enabling a thorough understanding of the characteristics of the assessment program, design choices made and (when applicable) student and teacher experiences. An interview guide was designed to direct the semi-structured interviews (see [Table 3](#)). Next to this, documentation was collected about the curriculum and assessments (e.g., assessment policy documents, study guides, etc.). The interviews were held face-to-face or by video call and were audio recorded. The interviews were transcribed verbatim. Both the recording and transcripts are stored in a secure research drive that can be accessed by the authors only. Informed consent was given by the participants on data collection, data storage and data use.

Data analysis

The semi-structured interviews and documents were analyzed thematically by the three authors collaboratively. A thick case description was developed for all nine programs, which was sent back to the participants for member checking. When necessary, these member check documents contained further clarifying questions, which were answered by the participants in the document itself, or in an additional short interview. The participants gave their consent on the final thick

TABLE 2 Overview of higher education programs and participants.

Case number	Domain	Participants	Level of education	Type of education	Experience with PA
1	Health	1 academic director, 1 teacher/assessor, 1 alumnus, 1 teacher/curriculum designer	University	Master	Partial implementation (workplace). Experience: several years
2	Health	1 teacher/curriculum designer	University	Bachelor	Partial implementation (workplace). Experience: several years
3	Paramedics	2 teacher/curriculum designers	Higher professional education	Bachelor	Partial implementation (workplace). Experience: several years
4	Teaching	2 teachers/trainers	Higher professional education	Teacher professionalization	Experience: several years
5	Teaching	1 teacher/curriculum designer	Higher professional education	Bachelor	Experience: working on design
6	Paramedics	1 team leader/curriculum designer, 3 teachers, 1 student	Higher professional education	Bachelor	Experience: several years
7	Communication	2 teachers/curriculum designers	Higher professional education	Bachelor	Experience: several years
8	Teaching	2 teachers/curriculum designers	Higher professional education	Master	Experience: complete design, first experiences in practice
9	Teaching	2 teachers/curriculum designers	Higher professional education	Teacher professionalization	Experience: first experiences in practice

TABLE 3 Interview guide for semi-structured interviews.

Context

Information about the profession and developments in the work field that informed curriculum and assessment design

Views on learning and assessment (e.g., assessment policy) that informed curriculum and assessment design

Why programmatic assessment

Design of the assessment program

Backbone of the program (learning outcomes, competences, ...)

Data-points, structure, and timing of data-points

Low, medium, high stakes decisions

Feedback and stimulating student learning through programmatic assessment

Teacher and student competence to enact programmatic assessment

Programmatic assessment in practice

Student experiences (e.g., of low-high stakes)

Teacher experiences

Design and implementation process

Timeline of design and implementation process

Positive and negative influences on design and implementation

description, which—in their opinion—correctly described their assessment program, design choices and experiences in practice.

Then, cross-case thematic analysis focused on similarities and differences between the cases with regard to design choices, adherence to PA principles and practical experiences. The analysis process can be characterized as a deductive approach (using the six PA principles as a template or sensitizing concepts), but also involving inductive analysis when identifying themes that characterize design choices and practical experiences. The analysis process was interpretive, with the goal of reaching an in-depth understanding of the cases, but inevitably involving the interpretation of the authors. Thematic analysis is systematic, but always subjective. We acknowledge that the data analysis of this study entails both interpretations and assumptions, based on the co-construction by the authors and the member checks by the interview participants (Watling and Lingard, 2012). All three authors thoroughly read the thick descriptions to get acquainted with the data and independently made an overview of recurring themes per principle of PA. These first findings were discussed in the research team. Based on this discussion, the first author (LB) made an overview of the themes within each of the six principles, supplemented with examples from the different cases. This overview was corrected and supplemented by the other authors (TS-M and CV) and again discussed in the research team. Finally, LB went back to the thick descriptions to check whether no information from the cases was missing, resulting in a final overview of the six principles with underlying themes and illustrating quotations from the thick descriptions. This overview was used as a basis for the description of the results. Finally, the results as described below were sent to the participants for member checking and consent

for publication. This research has been ethically approved by the Ethical Research Committee of the HAN University of Applied Sciences, approval number ECO 357.04/22.

Results

The following paragraphs are organized per principle of PA, which were taken as the starting point of the analysis. In these descriptions, we included the design choices made within this principle, the adherence (or not) of design choices with the theoretical principle, differences between the cases (when applicable) and student and teacher experiences reflecting how the design choices work out in practice (when applicable). To start with, two general findings are worth mentioning. First, the participants were aware of the different choices they had, as this participant describes: “*Programmatic Assessment sometimes gives you the impression there is a ready-made package with instructions, like a Billy-closet you buy at Ikea. That really is not how it works. It is not about the system itself, but about the ideas behind the system of programmatic assessment. You have to make it fit to your own context*” (C5). Second, all participants described their reasons for implementing PA as the wish to integrate assessment in students’ learning process and prevent student from hopping from one assessment to another, without taking along feedback. Also, many participants phrased the wish to stimulate student involvement and self-regulation. They all voiced a clear view of the learning processes they considered important and the work field developments that should guide curriculum development. For all cases, PA was not the starting point of the design process, but an approach to assessment that fits their views on what is important in student learning.

“*At that point in time, we developed a new policy strategy including our view on education. . . in which we paid a lot of attention to activating students, giving feedback, diversity, inclusion, and professional identity. . . [. . .] the principles of programmatic assessment, like the use of feedback to stimulate self-regulation, creating student ownership about their development, seemed to fit what we wanted to realize in our education*” (C9).

Programmatic assessment principle 1: A framework of learning outcomes makes up the backbone of the program, providing the basis for data-points and aggregation

In all cases, a “backbone” can be recognized in the assessment program, based on the learning outcomes (competences, skills, etc.) of the specific professional domain.

This backbone is used in the design process to specify data-points that fit this backbone, and to develop assessment rubrics and feedback forms. The participants stressed the importance of this backbone and why all teachers need to be familiar with and agree to this backbone: *“In a conventional setting all teachers take care of a piece of the curriculum, or a subject. The teacher determines what subject knowledge he includes—this is my subject and I know what students need. In PA you have to transcend this level and together, from your view of what a profession entails, determine how all subjects relate to this profession.”* (C4). This quotation also shows that teachers might find it difficult to let go of their own subject and their ownership of a specified part or module in the curriculum, an obstacle that was also voiced by other participants. In one of the cases (C5), the participants reported they found it very difficult to determine and agree upon this backbone, again partly because teachers were afraid their own subject would not be assessed or valued. As a consequence, they could not specify data-points and the implementation of PA was postponed.

A backbone in PA consists of a limited number (in our cases, four to six) of learning outcomes, described as competences, roles, aspects of a profession, themes or skills. It seems important not to include too many learning outcomes: *“We chose not to use the competences as a backbone for our program. We have 12 competences and that would risk fragmentation”* (C6). This participant describes the same: *“We had 5 segments [from a model used by the university] and 16 learning goals. We did not want the learning goals to become our focus point, because students could tick off goals they reached. Therefore we used the learning goals for the design of learning activities, but they are not recognizable in the assessment form. We assess on the level of the segments of the profession”* (C9).

Summarizing, all cases adhered to this principle and used a backbone as the starting point for further design choices on data-points. The design of the backbone itself is deemed complicated, but essential for further implementation, as was shown by C5 in which no agreement could be reached on the backbone. Design choices differ with regard to the number of learning outcomes, but a manageable number (4–6 in our cases) seems important. Design choices with regard to how saturation takes place were not reported, though this is stressed in this (theoretical) principle as well.

Programmatic assessment principle 2: A mix of different data-points is used to gain insight in student development

In all cases, a combination of different data-points is used to gain insight in student development and to make high-stakes decisions. Feedback is an essential part of all data-points. In our nine cases, a data-point can thus be described as an “artefact”—something a student does in practice, creates,

demonstrates, etc.—and this artefact is always combined with feedback that represents different perspectives on the quality of the student’s work (i.e., the artefact). Feedback is given by teachers, experts, clients, customers, internship supervisors, fellow-students, and also comprises of self-assessment. All participants describe how they purposefully designed data-points that would fit the learning processes they aimed to elicit and the learning outcomes to be attained at the end of the program. Reasons to design data-points were guided by didactical considerations and the decision to be made later on, based on the aggregation of the information the data-points provide. Didactical considerations are for example phrased as: *“You run the risk here that people say: we do not want any more summative tests, so now we call everything a data-point. But you have to make deliberate choices in your didactical design in order to make a good decision about student learning . . . Not every Kahoot quiz counts as a data-point. You do not need to document everything. Data-points are meaningful activities that tell you something about the students’ learning process”* (C4). The backbone of learning outcomes (principle 1) was found to be an important starting point to determine appropriate data-points, leading to a backwards design process. In one of the programs (C8), the curriculum developers took the learning outcomes as a starting point to choose several data-points students could work on, which resulted in seven types of data-points: practical assignment, position paper, presentation, video-selfie, assessment interview, progress test, and additional evidence to be chosen by the student.

The context in which PA is implemented also influenced the design of data-points. When PA is implemented in the context of workplace learning, the artefacts mainly comprise student performances in practice, like making a diagnosis or treating a patient (C1, C3). In these programs, multiple forms were developed to capture a student’s performance, like observation forms or forms to be filled out by patients and other persons involved. Some programs determined mandatory data-points. For example, in teacher education (C8, C9), students have to include video-taped lessons, as this was considered essential to be able to make a decision about a student’s teaching competence. In some programs, knowledge tests are considered important as vocational knowledge comprises an important part of students’ competence (C6, C7). Knowledge tests are explicitly used as data-points, as is explained by this participant, to make students aware of possible knowledge gaps and stimulate learning: *“Students do have to make the knowledge test, but they do not have to “pass” this test. No pass/fail decision is based on this test. Students get to know the percentage of questions they answered correctly. They put this in their portfolio. Besides these knowledge tests, students demonstrate knowledge in various practical products they make”* (C7).

Finally, the participants told they want to gradually increase students’ responsibility for their learning by letting students choose their own data-points. For example, first-year students

get a list of data-points or feedback to be collected (C6), while experienced teachers learning on the job choose their own data-points based on their urgent learning needs or experienced pride in trying out new things in their classrooms (C9). Design choices of all programs thus seem to adhere to the (theoretical) PA principle: data points are diverse, always involve feedback and are chosen based on didactical grounds and decision-making later on (connected to the backbone; principle 1). The actual choice for (types of) data-points differs per program, depending on contextual factors like the profession, the importance of knowledge and the student population.

Programmatic assessment principle 3: All data-points focus on optimizing feedback and do not involve pass/fail decisions

All nine cases adhered to this principle. In one program, however, students do have to score “as expected” on all data-points before they can enter the final assessment interview, which—strictly viewed—means that pass/fail decisions are connected to individual data-points (C7). The participants of this case tell about their design choices, the reasons for these choices and effects in practice: “*We deliberately chose to only let students enter the interview when their practical assignments are validated as “as expected” by the experts. The interviews are time-consuming. . . if a pass is considered impossible beforehand because the evidence is below-level, we wanted to protect teachers against needless work*” (C7). And: “*One thing we experience in practice. . . [tells about this design choice]. . . this causes friction. Some students get a “below expectations” in October. They have a problem, because they cannot enter the interview. Actually, we already make the decision at that point in time . . .*” (C7).

To enable students to use feedback, a series of data-points is needed in which competences or skills can be repeatedly practiced. For learning purposes, it is deemed important that negative feedback is not perceived as problematic, as this involves many learning opportunities. Students can take along feedback in next data-points and show improvement, as this participant explains: “*A portfolio can contain work that did not go very well. We stimulate students to put negative experiences in their portfolio, because this is an opportunity to set new learning goals, reflect on it, change behavior in practice and collect new evidence*” (C3).

In practice, students sometimes experience data-points to be high-stakes. This participant phrased students’ feelings as follows: “*Students thought “this information will be in my portfolio and will never get out anymore.” This requires training, support and guidance. I try to explain to students that they are assessed, but that the main point is to show how they take up possible critical feedback. That nothing happens if they show that they learned from feedback. On the contrary, that*

is what we want to see. We try to develop a feedback culture focused on learning” (C1). Stakes are in the eye of the beholder. Our cases show that for students, high stakes not only relate to pass/fail decisions, but also to experts or fellow-students possibly giving critical feedback on your work, or giving a presentation and possibly loose face. To lower students’ high-stakes feelings, the participants reported to increase students’ ownership. This is done by letting students fill out feedback forms first before asking others (C3), by letting students lead discussions with their mentors, set the agenda and determine their learning goals (C3), and by letting students invite feedback givers and asking them to reflect on the feedback (C8). Finally, intermediate reflections with mentors seem important to help students understand the meaning of the continuum of low-high stakes decisions: “*Imagine a student whose development is “below expectations” at the first intermediate evaluation. This student immediately realizes the second intermediate evaluation is meant to show development. Because this “below expectations” has no consequences yet, but does give the student a lot of information*”(C6). Summarizing, with one exception, all cases adhered to the principle that data-points do not involve pass/fail decisions and focus on feedback. In practice, however, students sometimes experience data-points to be high stakes and are afraid of negative feedback. Some programs therefore deliberately made design choices to diminish these high-stakes feelings by increasing students’ ownership and including medium stakes discussions with mentors.

Programmatic assessment principle 4: A continuous dialogue about data-points, feedback, and self-regulation

With regard to this fourth principle of PA, dialogues about data-points gathered so far are mostly designed by means of planned (medium-stakes) meetings between students and a teacher in the role of mentor/coach. These meetings can be individual (C1, C3), in groups of students (C6, C7), or planned incidentally when a student is recognized as at-risk (C2) and take place 2-weekly to twice a year. The goals of these meeting are described as follows: “*The reflective meetings are very important for the student’s learning process. We stress that these meetings are meant for reflection, as a formative moment.*” (C3). “*Students learn to make the connection between the practical products [i.e., data-points] they work on, and their long term development on the six competences. It requires extensive explanation to make this connection. Initially, students are mainly focused on current affairs. . . they have to learn to see the practical products as a means, to translate it to what it means for their development*” (C7).

Reflection on long-term development is thus a central goal of these meetings. Students are also stimulated to gradually take

ownership. For example, they are stimulated to ask for feedback (C6, C7), to plan sessions (C3), and to gather data-points themselves (C8, C9). This participant describes how student behavior changed as a result of PA: “After 3–4 months students go to other teachers to ask for feedback, they make appointments with one another, plan sessions. They are more independent, take the lead. We have teachers from other programs, and they noticed the same.” (C6). Teachers’ and students’ experiences on this principle seem largely positive: students and teachers change their mindset toward assessments, students show more self-regulation and actively seek feedback. Training for students and teachers seems imperative, however, to change existing mindsets toward assessment as a primarily summative endeavor. Also, teachers need to get used to their new roles as coaches and learn how to give feedback: “It requires a different mindset, especially when it comes to guidance. How do you monitor development, how explicitly do you express your opinion in feedback? You have to dare to say “that was not so good.” This is different from a student who fails a test at the end of the year.” (C1). Summarizing, this principle of PA was manifested in practice in all cases, by the design choice of intermediate meetings as the way to stimulate the dialogue about feedback and self-regulation. Different design choices were made to stimulate students to take ownership. Experiences in practice seem to show changed mindsets toward assessment, but teacher and student training seem essential with regard to their changing roles as mentor, feedback givers and feedback seekers.

Programmatic assessment principle 5: There is a proportional relationship between the number of data-points and the stakes of the decision

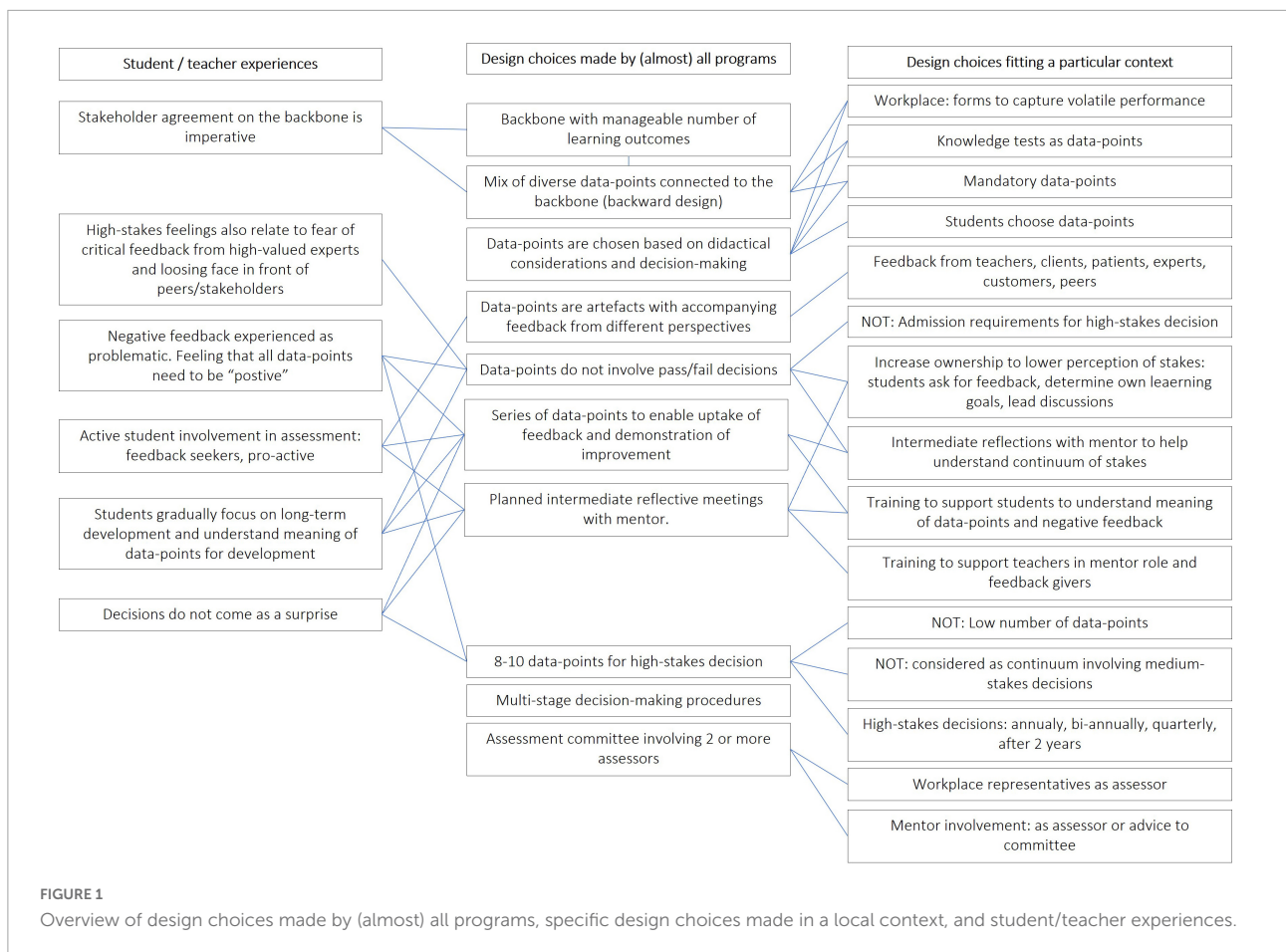
The number of data-points used for high-stakes decisions in our cases varies. The lowest number is three (C2), whereas other cases report about 8–10 data-points per learning outcome, and additionally report that student generally gather more data-points than required. High-stakes decisions are made mostly annually or bi-annually (C2, C3, C4, C6, C7), or after two full study years (C1) or quarterly (C8). In terms of credit points, 15–60 ECTS are awarded, whereas in other cases the decision is about a period of workplace learning. With regard to the number of data-points needed for high-stakes decision making, three data-points seem insufficient. The curriculum designers are looking for ways to include more data-points, but efficiency was a key consideration in the design: “The number of data-points is an area of tension. Now, just the three feedback sessions are mandatory, but not the training sessions, which also contain a lot of information. Students think so as well. We expect students to be present at the training sessions, though. These are not formal data points, but we do note when a student does not attend. But it is not feasible to take that into account [...]. The coaches

do report their experiences with students in a monitoring system [...] in our team meetings the experiences from other coaches are shared as well [...] that is how, next to the three feedback sessions, we try to incorporate other teachers’ informal experiences” (C2).

With regard to the number of data-points, an issue of design was whether data-points need to be “positive.” This topic already surfaced for PA principle 3 (for learning purposes), but again was a topic of discussion for this principle (for decision-making purposes). Again, the participants mentioned a considerable change in mindset toward assessment, both for students and teachers: “In the training, students learn it is no problem to get negative feedback, and that they should put this in their portfolio. That is how you learn. It requires a change in mindset. Not gathering feedback to show you master something, but gathering feedback to learn from it.” (C3). Due to the large number of data-points gathered during a long period of time, high-stakes decisions almost never come as a surprise. This is reported by almost all participants, for example voiced as follows: “The decision that remediation is necessary generally is not a surprise. In the intermediate meetings with the mentor this has been discussed already. The mentor helps the student to timely point out deficiencies” (C1). Summarizing, this principle of PA was mostly operationalized in terms of the number of data-points required for high-stakes decision making. The participants did not talk about a proportional relationship and—for example—the number of data-points needed for the intermediate stakes decisions (e.g., in the mentor meetings). The number of data-points for high stakes decision-making differs, due to practical constraints. A large number of data-points also mean that “mistakes” are not problematic and decisions do not come as a surprise.

Programmatic assessment principle 6: The higher the stakes of the decision, the more assessor expertise is needed

For high-stakes decisions, all cases in this study put in place multi-stage procedures in which more assessors are involved in decision-making in case of doubt. In almost all cases, an assessment committee makes the holistic high-stakes decision, based on the data-points in a portfolio. The assessment committee generally consists of two or more assessors, sometimes including work field professionals (C8). In one case, the mentor/coach is the primary decision-maker, but in case of doubt the student is discussed in a mentor/coach meeting, incorporating the opinion of several mentors/coaches in the final decision (C1). Pragmatic design choices thus have been made, lowering work pressure for the assessment committee. An often used multi-stage procedure involves two assessors who judge a student’s portfolio and make the decision, but in case of doubt send the portfolio to an additional



committee member, or put forward the portfolio for discussion in the committee meeting.

Members of the assessment committee are not fully independent of the student, that is, in many cases mentors/coaches are part of the assessment committee (C3, C6, C7, C8, C9) or send their advice about the student's progress as input to the committee (C6, C8). This could be a deliberate design choice: *"Information from the mentor/coach can be used as well. And for the mentor/coach it is useful to be part of the assessment committee and discuss how a student is doing. He can use that information for further guidance in the next semester"* (C7). Summarizing the findings with regard to this last principle of PA, the programs involved in this study seem to adhere to this principle. Design choices made regard multi-stage procedures, an assessment committee involving multiple assessors sometimes including work field professionals and mentors (who have diverse roles in decision-making). The participants mentioned design choices with regard to high-stakes decision-making only. They did not view this principle as involving a continuum of stakes, in which design choices could also involve the required assessor expertise at medium-stakes decisions.

Conclusion and discussion

The goal of this study was to explore: (1) what design choices are made by different programs that implemented PA, (2) to what extent these design choices adhere to the theoretical principles of PA, (3) how these design choices affect teachers and students' experience in practice, and (4) whether different design choices are made between the programs.

Design choices are viewed to be highly context-dependent (Carvalho and Goodyear, 2018; Bouw et al., 2021; Heeneman et al., 2021), and indeed a wide range of design choices was found across the nine programs showing a wide range of manifestations of PA in practice. Higher education institutes seem to use the principles of PA to make design choices that fit their particular students, views on learning, and workplace characteristics and developments. Figure 1 depicts our conclusions with regard to the design choices made by (almost) all programs included in this study, design choices made by some programs fitting their local context, and student/teacher experiences expressed with regard to the design choices (i.e., how these worked out in practice). It needs to be noted that Figure 1 depicts the interpretation of the

researchers and lines between the boxes do not represent causal relationships. Also, this study did and could not aim for saturation, in the sense that all possible design choices were identified.

Some design choices seem to be more “general,” as they were found in all nine programs. These design choices all adhered to the principles of PA. These design choices relate to the backbone of learning outcomes and the design of a mix of data-points related to this backbone. Data-points are designed to not involve pass/fail-decisions, always include feedback from different perspectives and are designed in a longitudinal way based on didactical considerations and later decision-making. All programs designed intermediate reflective meetings with a mentor. With regard to decision-making, 8–10 data-points are used as input for decision-making by a committee involving two or more assessors and multi-stage decision-making procedures.

Differences between the programs were found as well, and these context-specific design choices seem to represent alignments to professional domains, practical considerations and some choices that do not (completely) adhere to the principles of PA. For example, when PA is implemented during workplace learning—as in most health sciences education contexts—data-points are usually volatile and observations forms are used to capture student performance. In teacher education, video-taped lessons are mandatory data-points. Also, feedback givers represent different professional domains, such as clients, customers, experts, or peers. The number of assessors involved and the involvement of the mentor in decision-making differs per program. Although preference is often given to a committee with fully independent assessors (van der Vleuten et al., 2012), in many programs the mentor is part of the committee or otherwise involved in decision-making. This is done sometimes for pragmatic reasons (reducing workload, feasibility), in other cases this is a conscious design choice (the mentor/coach has a lot of information about the student that can help interpret student performance).

Some design choices warrant further discussion, as they do not (completely) adhere to the theoretical principles of PA, or seem to bring about negative student/teacher reactions. One program reported admission requirements for high-stakes decision-making, with concomitant student experiences of high-stakes decisions connected to single data-points [see also Baartman et al. (2022) for a more in-depth study]. Another program reported a low number of data-points, a concession made because of feasibility. With regard to data-points not involving pass/fail-decisions, several programs reported “high-stakes feelings,” in particular the fear of negative feedback. As reported in previous studies in PA, the continuum of stakes seems to be difficult to enact and understand for teachers and students (Bok et al., 2013; Heeneman et al., 2015; Harrison et al., 2016; Schut et al., 2018). The informants in this study seemed to be aware of this problem and made various design choices to lower the perception of stakes,

such as increasing students’ ownership, intermediate reflections and training. Torre et al. (2021) also reported a particularly wide range of implementations with regard to this principle of PA. However, most attention seems to be paid to the extremes of the continuum of stakes. For low-stakes, design choices align with the idea that feedback is the primary focus. For high-stakes, design choices focus on decisions made by multiple assessors collaboratively, who aggregate and interpret information from multiple data-points and substantiate their judgments [as reported by Oudkerk Pool et al. (2018)]. Design choices with regard to intermediate-stakes decisions seem to be less explicitly connected to a continuum of stakes. Intermediate reflective meetings are mentioned as very important for students to understand the meaning of data-points, to help students to gradually focus on long-term development and understand the value of (negative) feedback. This importance of intermediate meetings has been reported in other studies as well (e.g., Baartman et al., 2022). However, questions remain as to the number of data-points needed for meaningful intermediate decision-making, the timing of intermediate-decision-making, what kind of data-points and feedback are most helpful, and what enables students to take feedback on board and improve learning. The theoretical notion of feedback as a cyclic process involving dialogue and meaning-making as expressed in PA principle 4 (Carless et al., 2011; Boud and Molloy, 2013; Gulikers et al., 2021) needs to be further carved out in practical design choices. Here, recent research into feedback literacy of students and teachers (Boud and Dawson, 2021; Nicola-Richmond et al., 2021) might enhance the design choices made, and the implementation of PA in practice.

Finally, it needs to be noted that for participation in this study, programs were selected that (seemed to) understand and adhere to the theoretical principles of PA. There may be other programs that aimed to implement PA, but in which the implementation of PA was not successful because of contextual factors, cultural values, curriculum structure, or other barriers. Also, we might have missed programs that designed their assessments according to many of the principles of PA, without explicitly mentioning programmatic assessment as a starting point or approach to assessment. Future research could focus on important misconceptions of PA or design choices not fitting the PA principles. This could to inform practice on possibly “wrong” design choices, and broaden or ameliorate our theories on PA. Taking these limitations into account, this study contributes to our understanding of the operationalization of the PA principles in concrete design choices made by educational programs, aligned to their local context. We have seen that all programs made very conscious design choices, guided by their vision on learning, the backbone of learning outcomes, characteristics of the professional domain, and practical feasibility. Design choices indeed seem to be context-specific (Carvalho and Goodyear, 2018; Bouw et al., 2021; Heeneman et al., 2021), but more in-depth research is needed on how contextual factors like the

professional domain or student/teacher characteristics influence design choices made. The six theoretical principles of PA used as the starting point of this study seem to capture the essence of PA. Compared to the Ottawa consensus statement (Heeneman et al., 2021), we specifically emphasized the importance of a backbone of learning outcomes as the starting point of many design choices. Refinements might be needed with regard to the importance of intermediate-stakes decisions, which could be better phrased in the theoretical principles to emphasize its importance.

Data availability statement

The datasets presented in this article are not readily available because data are stored in a secure research drive. Participants have not been asked if the dataset can be shared with others.

Ethics statement

The studies involving human participants were reviewed and approved by Ethical Research Committee of the HAN University of Applied Sciences, approval number ECO 357.04/22. The patients/participants provided their written informed consent to participate in this study. Written informed

consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

LB and TS-M conducted the interviews. All authors analyzed the data, wrote the manuscript, and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Baartman, L. K. J., Baukema, H., and Prins, F. J. (2022). Exploring students' feedback seeking behavior in the context of programmatic assessment. *Assess. Eval. High. Educ.* 1–15. doi: 10.1080/02602938.2022.2100875
- Baartman, L. K. J., and Prins, F. J. (2018). Transparency or stimulating meaningfulness and self-regulation? A case study about a programmatic approach to transparency of assessment criteria. *Front. Educ.* 3:104. doi: 10.3389/feduc.2018.00104
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *High. Educ.* 32, 347–364. doi: 10.1007/BF00138871
- Bok, H. G. J., Teunissen, P. W., Favier, R. P., Rietbroek, N. J., Theyse, L. F. H., Brommer, H., et al. (2013). Programmatic assessment of competency-based workplace learning: When theory meets practice. *BMC Med. Educ.* 13:123. doi: 10.1186/1472-6920-13-123
- Boud, D., and Dawson, P. (2021). What feedback literate teachers do: An empirically-derived competency framework. *Assess. Eval. High. Educ.* doi: 10.1080/02602938.2021.1910928
- Boud, D., and Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assess. Eval. High. Educ.* 38, 698–712. doi: 10.1080/02602938.2012.691462
- Bouw, E., Zitter, I., and de Bruijn, E. (2021). Designable elements of integrative learning environments at the boundary of school and work: A multiple case study. *Learn. Environ. Res.* 24, 487–517. doi: 10.1007/s10984-020-09338-7
- Carless, D., Salter, D., Yang, M., and Lam, J. (2011). Developing sustainable feedback practices. *Stud. High. Educ.* 36, 395–407.
- Carvalho, L., and Goodyear, P. (2018). Design, learning networks and service innovation. *Des. Stud.* 55, 27–53. doi: 10.1016/j.destud.2017.09.003
- Cleland, J., MacLeod, A., and Ellaway, R. H. (2021). The curious case of case study research. *Med. Educ.* 55, 1131–1141. doi: 10.1111/medu.14544
- de Jong, L. H., Bok, H. G. J., Kremer, W. D. J., and van der Vleuten, C. P. M. (2019). Programmatic assessment: Can we provide evidence for saturation of information? *Med. Teach.* 41, 678–682. doi: 10.1080/0142159X.2018.1555369
- de Vos, M. E., Baartman, L. K. J., van der Vleuten, C. P. M., and de Bruijn, E. (2019). Exploring how educators at the workplace inform their judgement of students' professional performance. *J. Educ. Work* 32, 693–706. doi: 10.1080/13639080.2019.1696953
- Driessen, E. W., van Tartwijk, J., Govaerts, M. J. B., Teunissen, P., and van der Vleuten, C. (2012). The use of programmatic assessment in the clinical workplace: A Maastricht case report. *Med. Educ.* 34, 226–231. doi: 10.3109/0142159X.2012.652242
- Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *Am. Psychol.* 39, 193–202. doi: 10.1037/0003-066X.39.3.193
- Ginsburg, S., van der Vleuten, C. P. M., and Eva, K. W. (2017). The hidden value of narrative comments for assessment: A quantitative reliability analysis of qualitative data. *Acad. Med.* 92, 1617–1621. doi: 10.1097/ACM.0000000000001669
- Gulikers, J., Veugen, M., and Baartman, L. (2021). What are we really aiming for? Identifying concrete student behavior in co-regulatory formative assessment processes in the classroom. *Front. Educ.* 6:750281. doi: 10.3389/feduc.2021.750281
- Harrison, C. J., Könings, K. D., Dannefer, E. F., Schuwirth, L. W., Wass, V., and van der Vleuten, C. P. (2016). Factors influencing students' receptivity to formative feedback emerging from different assessment cultures. *Perspect. Med. Educ.* 5, 276–284. doi: 10.1007/s40037-016-0297-x
- Heeneman, S., and de Grave, W. (2017). Tensions in mentoring medical students toward self-directed and reflective learning in a longitudinal portfolio-based mentoring system—an activity theory analysis. *Med. Teach.* 39, 368–376. doi: 10.1080/0142159X.2017.1286308

- Heeneman, S., de Jong, L. H., Dawson, L. J., Wilkinson, T. J., Ryan, A., Tait, G. R., et al. (2021). Ottawa 2020 consensus statement for programmatic assessment–1. Agreement on the principles. *Med. Teach.* 43, 1139–1148. doi: 10.1080/0142159X.2021.1957088
- Heeneman, S., Pool, A. O., Schuwirth, L. W. T., Vleuten, C. P. M., and Driessen, E. W. (2015). The impact of programmatic assessment on student learning: Theory versus practice. *Med. Educ.* 49, 487–498. doi: 10.1111/medu.12645
- Imanipour, M., and Jalili, M. (2016). Development of a comprehensive clinical performance assessment system for nursing students: A programmatic approach. *Jpn. J. Nurs. Sci.* 13, 46–54. doi: 10.1111/jjns.12085
- Jamieson, J., Jenkins, G., Beatty, S., and Palermo, C. (2017). Designing programmes of assessment: A participatory approach. *Med. Teach.* 39, 1182–1188. doi: 10.1080/0142159X.2017.1355447
- Jessop, T., and Tomas, C. (2017). The implications of programme assessment patterns for student learning. *Assess. Eval. High. Educ.* 42, 990–999.
- Kickert, R., Meeuwisse, M., Stegers-Jager, K. M., Prinzie, P., and Arends, L. R. (2021). Curricular fit perspective on motivation in higher education. *High. Educ.* 83, 729–745. doi: 10.1007/s10734-021-00699-3
- Nicola-Richmond, K., Tai, J., and Dawson, P. (2022). Students' feedback literacy in workplace integrated learning: How prepared are they? *Innov. Educ. Teach. Int.* 1–11. doi: 10.1080/14703297.2021.2013289
- Oudkerk Pool, A., Govaerts, M. J. B., Jaarsma, D. A. D. C., and Driessen, E. W. (2018). From aggregation to interpretation: How assessors judge complex data in a competency-based portfolio. *Adv. Health Sci. Educ.* 23, 275–287. doi: 10.1007/s10459-017-9793-y
- Price, M., Carroll, J., O'Donovan, B., and Rust, C. (2011). If I was going there I wouldn't start from here: A critical commentary on current assessment practice. *Assess. Eval. High. Educ.* 36, 479–492. doi: 10.1080/02602930903512883
- Ramani, S., Könings, K. D., Ginsburg, S., and van der Vleuten, C. P. M. (2020). Relationships as the backbone of feedback: Exploring preceptor and resident perceptions of their behaviors during feedback conversations. *Acad. Med.* 95, 1073–1081. doi: 10.1097/ACM.0000000000002971
- Sadler, D. R. (2007). Perils in the meticulous specification of goals and assessment criteria. *Assess. Educ. Princ. Policy Pract.* 14, 387–392. doi: 10.1080/09695940701592097
- Schut, S., Driessen, E., van Tartwijk, J., van der Vleuten, C., and Heeneman, S. (2018). Stakes in the eye of the beholder: An international study of learners' perceptions within programmatic assessment. *Med. Educ.* 52, 654–663. doi: 10.1111/medu.13532
- Schut, S., Maggio, L. A., Heeneman, S., van Tartwijk, J., van der Vleuten, C., and Driessen, E. (2020a). Where the rubber meets the road — an integrative review of programmatic assessment in health care professions education. *Perspect. Med. Educ.* 10, 6–13. doi: 10.1007/s40037-020-00625-w
- Schut, S., van Tartwijk, J., Driessen, E., van der Vleuten, C., and Heeneman, S. (2020b). Understanding the influence of teacher–learner relationships on learners' assessment perception. *Adv. Health Sci. Educ.* 25, 441–456. doi: 10.1007/s10459-019-09935-z
- Torre, D., Rice, N. E., Ryan, A., Bok, H., Dawson, L. J., Bierer, B., et al. (2021). Ottawa 2020 consensus statements for programmatic assessment–2. Implementation and practice. *Med. Teach.* 43, 1149–1160.
- Torre, D. M., Schuwirth, L. W. T., and van der Vleuten, C. P. M. (2020). Theoretical considerations on programmatic assessment. *Med. Teach.* 42, 213–220. doi: 10.1080/0142159X.2019.1672863
- van den Akker, J. (2004). "Curriculum perspectives: An introduction," in *Curriculum landscapes and trends*, eds J. van den Akker, W. Kuiper, and U. Hamever (Dordrecht: Kluwer Academic), 1–10.
- van der Vleuten, C. P. M. (1996). The assessment of professional competence: Developments, research and practical implications. *Adv. Health Sci. Educ.* 1, 41–67. doi: 10.1007/BF00596229
- van der Vleuten, C. P. M., Schuwirth, L. W. T., Scheele, F., Driessen, E. W., and Hodges, B. (2010). The assessment of professional competence: Building blocks for theory development. *Best Pract. Res. Clin. Obstet. Gynaecol.* 24, 703–719. doi: 10.1016/j.bpobgyn.2010.04.001
- van der Vleuten, C. P. M., and Schuwirth, L. W. T. (2005). Assessing professional competence: From methods to programmes. *Med. Educ.* 39, 309–317. doi: 10.1111/j.1365-2929.2005.02094.x
- van der Vleuten, C. P. M., Schuwirth, L. W. T., Driessen, E. W., Dijkstra, J., Tigelaar, D., Baartman, L. K. J., et al. (2012). A model for programmatic assessment fit for purpose. *Med. Teach.* 34, 205–214. doi: 10.3109/0142159X.2012.652239
- Watling, C. J., and Lingard, L. (2012). Grounded theory in medical education research: AMEE Guide No. 70. *Med. Teach.* 34, 850–861.
- Wilkinson, T. J., Tweed, M. J., Egan, T. G., Ali, A. N., McKenzie, J. M., Moore, M., et al. (2011). Joining the dots: Conditional pass and programmatic assessment enhances recognition of problems with professionalism and factors hampering student progress. *BMC Med. Educ.* 11:29. doi: 10.1186/1472-6920-11-29
- Yazan, B. (2015). Three approaches to case study methods in education: Yin, Merriam, and Stake. *Qual. Rep.* 20, 134–152.