



Kent Academic Repository

Baartman, Liesbeth K. J. and Quinlan, Kathleen M. (2023) *Assessment and feedback in higher education reimagined: using programmatic assessment to transform higher education*. Perspectives: Policy and Practice in Higher Education, 28 (2). pp. 57-67. ISSN 1360-3108.

Downloaded from

<https://kar.kent.ac.uk/104168/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.1080/13603108.2023.2283118>

This document version

Publisher pdf

DOI for this version

Licence for this version

CC BY-NC-ND (Attribution-NonCommercial-NoDerivatives)

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in **Title of Journal**, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).



Assessment and feedback in higher education reimagined: using programmatic assessment to transform higher education

Liesbeth K. J. Baartman & Kathleen M. Quinlan

To cite this article: Liesbeth K. J. Baartman & Kathleen M. Quinlan (24 Nov 2023): Assessment and feedback in higher education reimagined: using programmatic assessment to transform higher education, *Perspectives: Policy and Practice in Higher Education*, DOI: [10.1080/13603108.2023.2283118](https://doi.org/10.1080/13603108.2023.2283118)

To link to this article: <https://doi.org/10.1080/13603108.2023.2283118>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 24 Nov 2023.



[Submit your article to this journal](#)



Article views: 795



[View related articles](#)



[View Crossmark data](#)

Assessment and feedback in higher education reimagined: using programmatic assessment to transform higher education

Liesbeth K. J. Baartman ^a and Kathleen M. Quinlan ^b

^aResearch Group Vocational Education, Utrecht University of Applied Sciences, Utrecht, the Netherlands; ^bCentre for the Study of Higher Education, University of Kent, Canterbury, UK

ABSTRACT

We argue that assessment and feedback practices in higher education need to be transformed to better address three purposes: *promoting* learning, *assuring* assessment rigour, and *communicating* students' employability. To address shortcomings in the current assessment and feedback culture, we propose programmatic assessment (PA), a new approach to assessment developed initially in medical education and now applied to a range of other professional fields. We outline eight recommendations that synthesise the key principles underpinning PA. Then, drawing on experience with PA in various professional fields in the Netherlands, we describe and illustrate four action steps for programme teams to take to implement programmatic assessment. We highlight implications of such a shift for leaders and professional services staff before concluding that PA can transform higher education by creating a more productive culture of assessment and feedback.

ARTICLE HISTORY

Received 15 April 2023
Accepted 7 November 2023

KEYWORDS

Assessment; programmatic assessment; feedback; higher education; employability

Introduction: the need for transformation

Assessment and feedback practices are among the most powerful influences on students' learning in higher education (HE). Unfortunately, higher education students around the world often experience a testing culture, involving many summative assessments and a tendency to teach and learn to tests (Jessop and Tomas 2017). Students are traditionally assessed after each module, leading to a pass/fail-decision about attainment on that module. If the student passes all modules, a degree is awarded. Assessment of learning is emphasised, viewing assessment only as the 'means by which we assure and express academic standards' (Elkington 2020, 5).

In many countries, there is growing recognition that we also need to design assessment *for* learning (Stobart 2008; Black and Wiliam 2018). From this perspective, learners engage in assessment tasks that generate data (information) that can be used as evidence by students, peers, teaching staff (i.e. academics in higher education), or employers to support further learning. This broadened conception of assessment 'puts learners' activities at the centre, rather than the academics who are expressing and assuring academic standards' (Pitt and Quinlan 2022, 13). Assessment *for* learning (AfL) emphasises students' understanding of intended learning outcomes and assessment criteria and completion of meaningful assessment tasks that are aligned with those learning outcomes. In this

approach, assessment and feedback systems are designed to ensure that students take ownership of their own learning, engage with peers as resources in that learning, and seek out, make sense of, and use feedback from various sources to promote their learning (Carless 2015; Henderson et al. 2019; Nicol 2021; Winstone and Boud 2019).

While evidence is mounting that implementing core principles of AfL supports student satisfaction, engagement, and learning within single modules, a wide-ranging review of Anglophone literature originating from 71 countries found that little attention has been paid to changing the culture of assessment and feedback across whole programmes (Pitt and Quinlan 2022). Instead, AfL tends to be implemented by lone innovators within single, isolated modules. Unfortunately, their impact is limited within the traditional, modularised summative approach to assessment which inadvertently incentivises a raft of undesirable learning behaviours. For example, in modularised systems, students tend to focus on short-term extrinsic rewards (module grades) (Harland et al. 2015), treat learning in one module as separate from other modules (Kickert et al. 2021; Sadler 2007), and ignore feedback on end-of-module assessments (Harrison et al. 2016).

Within the traditional assessment system, using assessment information to communicate students' accomplishments to future employers is often

CONTACT Kathleen M. Quinlan  K.M.Quinlan@kent.ac.uk

overlooked. The current system typically relies on aggregated marks within modules, further averaged across modules for a single degree classification or grade point average or class rank. It communicates how well students performed across assessments within a module but does not provide information about specific competences or relative strengths, weaknesses, or the distinctive skills or attributes an individual student may have (Jorre de St Jorre, Boud, and Johnson 2021). Although the UK's Higher Education Achievement Report (HEAR) (Higher Education Academy 2015) provides a richer picture of students' achievements by incorporating extra-curricular activities and awards, it still tends to report academic achievements by marks in modules rather than reporting on underlying competences sought by employers. This misalignment between employers' needs and academic assessment systems that only list marks by module shifts the burden to careers services to educate students about how to identify, build, and represent their competences to employers.

Instead, we propose an assessment system that can serve all three of these purposes of assessment: *promoting* learning, *assuring* assessment rigour, and *communicating* employability.¹ Such a system should leverage assessment and feedback to promote students' responsibility for progress toward intended programme-level learning outcomes. It needs to uphold standards of validity and reliability across the entire course. And it should integrate and aggregate assessment information in a meaningful way to represent unfolding competences to students themselves, peers, academics, and employers.

Innovative idea: programmatic assessment

Programmatic assessment (PA) is a new approach to assessment developed initially in medical education (Van der Vleuten et al. 2012; Van der Vleuten et al. 2010) and now applied to a range of other professional fields, particularly in the Netherlands (Baartman, Baukema, and Prins 2022; Baartman, van Schilt-Mol, and van der Vleuten 2022). Although PA has been used primarily to date in professional education, it can also be adapted to a wide range of disciplines in higher education, so long as intended learning outcomes are clearly identified. PA entails a fundamental paradigm shift in our approach to both assessment *for* learning and assessment *of* learning. Founded on both learning theories (Torre, Schuwirth, and van der Vleuten 2020) and theoretical frameworks about validity and reliability (Van der Vleuten et al. 2010), it addresses both learning enhancement and assessment rigour agendas outlined above. A growing body of evidence supports the value of PA, demonstrating that it catalyses learning and enables robust decision-making (e.g. Driessen et al. 2012; Imanipour and Jalili 2016;

Schut et al. 2021). In PA, vertical integration of competences across programmes also aligns education with employability. Various consensus statements and seminal papers have articulated key principles of PA (Heeneman et al. 2021; Torre et al. 2021; Baartman, van Schilt-Mol, and van der Vleuten 2022) which we synthesise under eight key recommendations in this section. In the subsequent section, we translate these recommendations into four action steps with examples from practice in applied fields.

Detail intended learning outcomes in a competences framework

First, PA is based upon an agreed competences framework that makes up the 'backbone' of the programme. This backbone comprises a set of competences or complex skills that are specified at different levels that students must attain by the end of each phase of the programme (e.g. half a year or a year) (Baartman, van Schilt-Mol, and van der Vleuten 2022). Assessment methods, and eventually teaching methods, are then aligned with this competences framework (Biggs 1996). This competence framework can be built on whatever learning outcomes are set by a programme, whether those are aligned to specific professional societies or accreditation requirements, to general lists of skills sought by employers across sectors, to competencies related to sustainability, to general graduate attributes, to quality assurance benchmarks for the discipline, to lists of outcomes expected of liberal arts graduates, or any other set of disciplinary, institutional or societal expectations of graduates that a programme endorses.

Distinguish between assessment information and uses of that information

Crucially, PA makes a distinction between assessment information and what academics, students, and quality assurance staff do with this information. All assessment methods generate information about learners' progress. Assessment *for* learning means academics and students use assessment information to promote learning of the expected competences. Together, they diagnose student progress, make meaning of feedback in dialogue, and determine the next step in learning. Assessment *of* learning means academics (as assessors) use assessment information to decide on what has been learned and assure academic standards. Quality assurance officers can use assessment information directly as part of ensuring the quality of education. Finally, future employers use assessment information to judge the capabilities, strengths, and weaknesses of students they may want to employ.

Create a continuum of assessment 'stakes'

PA's distinction between *collecting* assessment information and *using* this information for multiple purposes replaces the traditional binary divide between formative and summative with a continuum of stakes, as illustrated in Figure 1. These range from low-stakes decisions with few consequences for students to high-stakes decisions with potentially great consequences. An example of a low-stakes decision is an academic's or work-based supervisor's feedback on the quality of the student's work-in-progress with suggestions for how the student could revise the work, continue practicing, or use the feedback in future assignments. Examples of high-stakes decisions are using assessment information to decide on whether a student can progress to another year or be awarded a particular degree classification. Within this continuum, PA also includes medium-stakes decisions that focus on identifying strengths and weaknesses in students' development, identifying areas for improvement, and setting new learning goals to work on.

Collect information about competences at multiple points longitudinally

Building on the idea of a continuum of stakes, the key innovation in PA is the longitudinal collection of assessment information about student learning along the backbone competence framework. Research shows that any single assessment – regardless of the assessment method – provides insufficient evidence for robust decision-making about learner progress on complex competences (Van der Vleuten et al. 2010). Therefore, as shown in Figure 1, high-stakes decisions

should be based on a multitude of assessments (called data points) that are aggregated and combined into a holistic judgement (Wilkinson et al. 2011). Data points are collected throughout the learning process and can comprise a multitude of forms of evidence about student learning such as knowledge tests, practical assignments, and observations in practice with feedback gathered from academics, (workplace) experts, peers, or clients. Single data points never involve pass/fail decisions and thus are used for low-stakes decisions only. They are aimed at generating useful feedback for further learning. High-stakes decisions are holistic judgements based on the combination, aggregation, and saturation of information from a large variety of data points that offer both quantitative and qualitative information about student progress.

Ensure proportionality between the stakes of the decision and the number of data points

There is a proportional relationship between the stakes of the decision and the number of data points needed for decision-making. As depicted in Figure 1, a single assessment is like just one pixel in a photograph. In HE, we usually make decisions about passing or failing a module based on one or perhaps two or three of these pixels. However, these single pixels provide an incomplete picture of students' capabilities. When more pixels (or data points in PA) are added, the picture of the student becomes sharper. At the left side of Figure 1, low-stakes decisions are made, based on a single or a small number of data points. These decisions could be seen as feedback, in which others (or the student) make inferences about the work of the student in the form of a momentary decision or

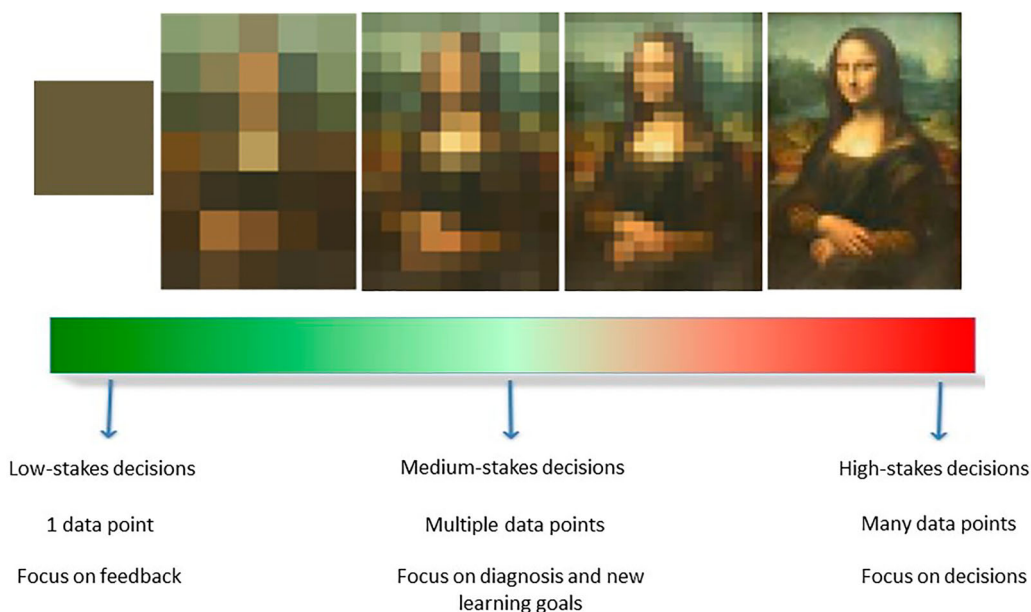


Figure 1. A continuum of stakes in programmatic assessment.

value judgment (De Vos et al. 2022) which the student subsequently uses to support learning. At the right side of Figure 1, those value judgements (feedback information) from various stakeholders are aggregated into a complex picture that enables panels of educators to make high-stakes decisions.

Support students to make meaning of data points and use feedback

Students usually collect and store their data points in an (electronic) portfolio, including both the student's work – the artefacts – and the feedback from others (e.g. academics, peers, workplace professionals) who judged the quality of these artefacts. Remember that none of these data points involve high-stakes decisions; in other words, students do not have to pass these assessments. A portfolio can thus contain 'bad' data points when performance was unsatisfactory at that moment and for that assignment. With the help of academics, students analyse and triangulate all information (data points) they have gathered, identifying strengths and weaknesses, and determining next steps in learning in relation to the intended learning outcomes. Feedback is not seen as simply comments on students' work or 'dangling data' as Sadler (1989) called it, but as a cyclic process that needs active and continuing student engagement (Boud and Molloy 2013). Dialogue is essential, as learning is seen as a social activity involving shared meaning making. Through this dialogue, PA stimulates students' self-regulated learning and evaluative judgement about their own progress (Schut et al. 2021). Fundamental to PA is the idea of meaning-making and uptake of feedback, which is central to recent literature on feedback literacy (e.g. Carless and Boud 2018; Molloy, Boud,

and Henderson 2020). For feedback processes to work, students need to appreciate feedback, be able to make judgements, manage emotions, and act on feedback. Feedback needs a landing space for students to be able to act on it. In PA, this landing space is ensured by the sequence of data points, connected by the backbone. Lastly, being able to judge, represent, and communicate about their own capabilities is a key skill students will need throughout their careers (Healy 2023). In PA, career development and programme assessment are typically aligned through a career-relevant competence framework. Unlike traditional assessments that aggregate horizontally across assessments (Figure 2), PA facilitates students' employability by focusing on aggregating evidence about competence development vertically throughout the programme (Figure 3), helping students to recognise and value their progress and developing their evaluative judgement.

Students and assessors aggregate information vertically about competences, rather than horizontally across assessments

In PA, decision-making (whether low-stakes or high-stakes) is based on meaningful aggregation of assessment information about the intended competences. The programme's competence framework provides the basis for both the collection of data points and the judgement and aggregation of information towards decision-making. Figure 2 depicts how aggregation of assessment information towards a judgement evolves in the old assessment paradigm versus in the new paradigm of PA in Figure 3. In Figure 2, assessment information is aggregated horizontally. Information about different learning outcomes/

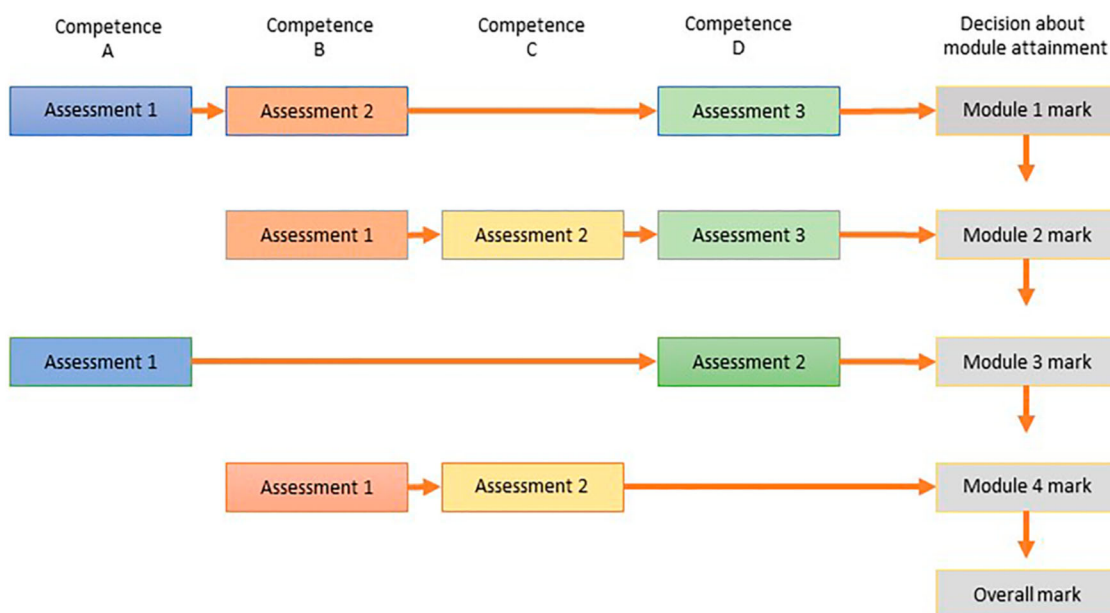


Figure 2. Horizontal aggregation of competence information in the old assessment paradigm.

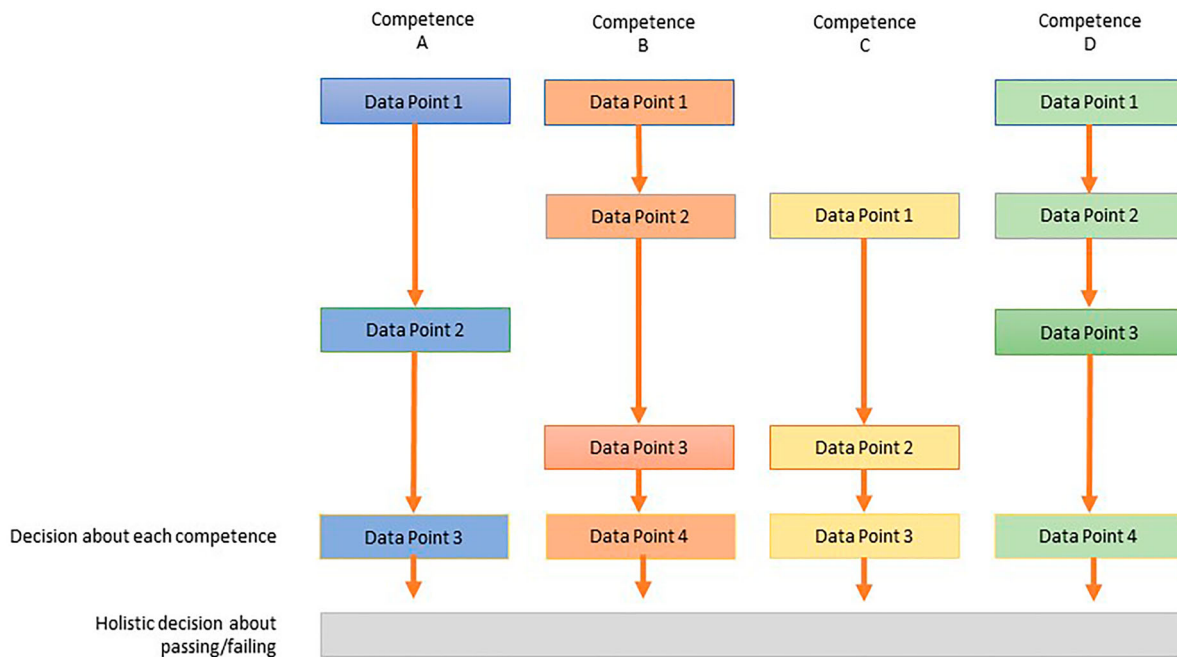


Figure 3. Vertical integration of competence information in PA.

competences is aggregated towards a judgement on attainment of an assessment or module. Thus, in the final mark, student performance on these different learning outcomes is obscured (Jorre de St Jorre, Boud, and Johnson 2021). A student who passes a module could perform well on competence A, but badly on competence B. In PA, information is instead aggregated vertically. Assessment information about a single learning outcome/competence is combined meaningfully from different data points (Figure 3). Although data points usually contain information about student progress on multiple competences, relevant information is extracted and aggregated to reach judgement on competence development instead of on ‘passing an assessment’. Thus, decision-making follows the programme’s competence framework, meaningfully aggregating information from data points towards a decision about each competence.

Involve multiple assessors reviewing many data points to assure rigour in high-stakes decision-making

In PA, decisions are based on information-rich data points, including both qualitative and quantitative information on student competences. As quantitative aggregation is therefore impossible, decision-making always involves human judgement, which brings the risk of bias. In PA, bias is first overcome by large sampling and triangulation (Van der Vleuten et al. 2012), as noted above. Multiple sources of evidence (i.e. data points) are sampled across a prolonged timeframe over different contexts, including many quality

judgements (i.e. feedback) by relevant stakeholders. Second, high-stakes decisions should not be made individually. Bias is overcome by ensuring expert committees make high-stakes decisions, based on clear group decision-making procedures (De Jong et al. 2019). Research shows that when making high-stakes decisions, although assessors use different, unique approaches to reviewing and judging assessment information (the data points in the portfolio), they eventually reach the same holistic judgements (Oudkerk Pool et al. 2018). Comparison across assessors mitigates subjective judgements and helps to build shared mental models for decision-making on complex constructs like competences. For most students, the information and progress is clear, and decision-making by the committee is straightforward and uncontested. A small subset of students may be on the borderline, though. In these cases, the committee deliberates, weighs information, and continues discussion until consensus is reached. Because saturation of information is a key quality criterion (De Jong et al. 2019), a committee can also decide that insufficient information (i.e. data points) is available to reach a robust decision. More time to collect data points and the inclusion of more perspectives (i.e. feedback from diverse others) is necessary before a high-stakes decision can be made.

How to make PA happen in practice

Although there is consensus about the theoretical principles that underpin the design of PA, various design choices can be made. Because curriculum design choices are highly context-specific (Heeneman

et al. 2021), they must be made by individual programme designers, such as programme leaders in consultation with local quality enhancement and assurance staff. Thus, PA can be described as a design issue. Acknowledging that there are different approaches a team could take, we suggest a backward design approach that starts with defining intended learning outcomes and working backwards through assessment and feedback design to teaching and learning approaches. In this section, we present a step-by-step guide to the key design decisions that programmes need to make to implement PA. Although we present these key design decisions as a step-by-step approach, we want to emphasise that actual design processes usually take place in an iterative and often unpredictable way.

Our guidance is based on existing examples of how PA has been designed and implemented. For example, in their review of PA practices in health sciences education, Torre et al. (2021) showed how educational programmes make different design choices, appropriate to their own students, academics, culture, and practicalities. In the Netherlands, the ideas of PA are rapidly being embraced by a wide range of HE programmes, within and beyond health sciences education. A professional learning community was established three years ago, led by the first author of this manuscript, in which academics, curriculum developers, examination board members, and assessment experts from 40 different programmes currently participate. They collaboratively share, reflect on, and improve their PA practices.

Drawing on the experiences of this professional learning community, the first author and colleagues published a Dutch book with nine worked-out examples of PA representing a variety of disciplines including teacher education, communication sciences, and occupational therapy (Baartman, van Schilt-Mol, and van der Vleuten 2020). The current 2022–2023 professional learning community represents various disciplines including business administration, information and communications technology (ICT), dietetics, languages, and biology. From these practical experiences, we have a growing understanding of what works during the design and implementation process, what can hinder the implementation, and how challenges can be addressed (Baartman, van Schilt-Mol, and van der Vleuten 2022). Some examples of how PA has been designed and works in practice are shared here, based on experiences in this professional learning community. These examples detail how to apply the theoretical principles of PA outlined above. We illustrate four key steps: (1) agreeing the backbone of learning outcomes, (2) designing a mix of data points to gain insight into student development, (3) using data points for feedback and learning, and (4) using data points for high-stakes decision-making.

Design of learning and teaching activities and resources would follow but are beyond the scope of this essay.

Step 1: agree a backbone of learning outcomes as the basis for data points and aggregation of information for decision-making

The PA design process usually starts with agreeing the learning outcomes students must demonstrate to get a degree. These learning outcomes can be described in terms of competences, core tasks of the profession, or different roles graduates play. Most programmes have five to eight learning outcomes. It is important to use a limited number of learning outcomes to prevent fragmentation. One dietetics programme, for example, includes seven key competences: food and nutrition expertise, communication, research, working together, entrepreneurship and marketing, management, and professionalism, subdivided into 25 learning outcomes demonstrated at three levels representing year 1 (introductory level), years 2 and 3 (intermediate level), and year 4 (ready for practice). The backbone is used to determine data points that align with these learning outcomes, and to develop rubrics and feedback forms. Often, the learning outcomes and different levels are described in a holistic rubric used throughout the entire programme to judge student work and give feedback on data points. These rubrics help students to focus on their long-term development across the entire programme.

Agreeing this backbone is not always straightforward. In one programme, planning was derailed when participating designers and academic teaching staff could not come to a shared understanding of the backbone. Academics stressed the importance of different separate subjects and were afraid their own subject would not be recognisable in the holistic assessment. They were afraid of losing ownership of their own subject and assessments and concerned that students did not have to pass their tests. This defending of turf hindered a common focus on programme-level learning outcomes and the further design of the data points. Other programmes have overcome this barrier by involving academics from the start of the (re)design process or by gradually redesigning existing modules into larger ones in which different subject areas are integrated.

Step 2: design a mix of data points to gain insight into student development

The next step typically involves determining the data points to be collected throughout a period of learning. First, these data points should align with the backbone to enable saturation and robust decision-making at the end of a given period of learning

(e.g. half a year or a year). Second, the diversity of data points should generate meaningful feedback and enable students to practice and develop. In effect, data points are a meaningful sequence of assignments or tasks that are connected to the backbone and build on each other to enable the uptake of feedback.

An example from a business administration course shows how design teams developed data points. As we recommend here, they used a backward design process starting from the decision they wanted to make at the end of the semester. Their backbone consists of four core tasks: designing and developing a value-case, (re)designing business processes, designing and facilitating collaboration in organisations, and personal professional development. Data points consist of professional artefacts made by students, together with feedback, self-assessments, and each student's reflections. Examples of data points are a written briefing about the feasibility, desirability and sustainability of a new product or service; a flow chart visual of a business process; and the description and enactment of conditions for collaboration in a project team. At the start of the programme, data points are fixed and structured. For example, all students might complete the flow chart assignment by a particular date. Progressing through the programme, students increasingly take on more responsibility for choosing data points that fit the learning outcomes, their specialisation, and the professional they want to become.

Data points differ per programme. When PA is implemented in the context of workplace learning, artefacts mainly comprise student performance in practice such as making a diagnosis, talking to clients, or giving a presentation for a business team. In these programmes, the artefact itself may happen at any time while at work and is ephemeral. Thus, feedback forms are used to capture different judgements (e.g. of patients, colleagues, or customers) of student performance in the moment. Some programmes explicitly include knowledge tests as data points. Importantly, students – consistent with the principles of PA – do not have to pass each of these knowledge tests. Tests are used as data points about students' competences, along with other data points in which knowledge can be demonstrated.

Although instruction might still occur in modular form with academics leading a specified unit of instruction, data points are designed and built *across* all the modules students are experiencing in a given block of time (e.g. a semester or a year). Thus there is an emphasis on integration across (former) modules, similar to existing synoptic assessments that bridge two or more modules (Constantinou 2020), though synoptic assessment does not necessarily involve the longitudinal collection of data points.

Step 3: use data points for feedback and learning

In PA, single data points occur throughout the teaching and learning process, with students receiving feedback on their performance that supports their learning. These single data points, as explained above, should never involve pass/fail decisions. In practice, we do see examples of programmes in which high-stakes decisions are made based on single data points, which we find has negative effects on student learning. For example, in a communication sciences programme in which students had to achieve an 'as expected' for each data point before they were allowed to enter the end-of-year interview, students reported stress, feeling that making mistakes was not allowed, and the perception that data points were 'being tested' rather than supporting learning (Baartman, Baukema, and Prins 2022).

It can be difficult for academics to let go of the idea that students must pass each of their assessments or even 'their' module. However, the robustness of the decision-making at Step 4 assures that students must eventually demonstrate the programme learning outcomes even if they do not do so on a given data point. In this way, PA is similar to the traditional Oxbridge approach in which tutors interact with students throughout the year, but marks are only awarded at the end of a year (Horn 2013). PA differs from the Oxbridge approach, though, in that students are collecting a portfolio of data points throughout the year for subsequent decision-making, rather than sitting large, timed exams at the end, which also produce only a single, albeit integrated, data point.

In all programmes, students collect feedback from diverse perspectives on their data points. In an ICT programme, students work on about 10 data points during a semester, after which a high-stakes decision is made (30 credits). The data points are designed to ensure that students can use the feedback they get on the first data point in the second one, and so on. Just as ICT professionals do, students work together in small groups in short-cycle 2-to-3-week assignments, often for real customers. Students use feedback forms to get oral narrative feedback, make notes of the feedback, organise it in their portfolio, and ask the feedback giver to check the correctness of their notes. Halfway through the semester, students write a reflective note in which they reflect on the questions 'how would I describe my development on the competences, looking at my data points and the feedback I received?' and 'What am I going to do next to further improve?'. Students start the next cycle with a learning story, in which they might describe how they want to improve on a programming language or working together. In this next short-cycle assignment they are assigned a role that

allows them to work on that competency, such as the role of group leader or taking care of the communication with the customers. Student group work is typically supported by instruction, especially in earlier stages of students' programmes.

In PA, students are guided in interpreting feedback and making sense of judgements from diverse feedback givers, whose opinions might differ. We find that both students and academics must get used to their new roles in PA. Students must get used to making mistakes, seeking feedback, and gradually taking ownership for their learning. In the ICT programme described above, the academics strive to create a feedback culture, but they also acknowledge that students need more guidance than they thought, for example in connecting the information in their data points to the competences and their longer-term development along the programme backbone.

Medium-stakes decisions, typically occurring through dialogues between students and academics, seem especially important for students to focus on long-term competence development and to grasp the meaning of data points. A paramedics programme, for example, works with low-, medium-, and high-stakes decisions. The medium-stakes decisions take place in dialogues between the student and a mentor, discussing at least three data points. The student prepares for the meeting by filling out a self-assessment rubric, sets the agenda for the meeting, and leads the discussion. This expectation stimulates ownership and lowers the perceived stakes. When discussing data points and feedback, students come to realise that negative feedback indeed has no immediate consequences in terms of passing or failing. Negative feedback is discussed, compared to feedback given by others, and new learning goals are set for the next period of learning. On the other hand, mentors report that it is important to also give students an impression of progress towards the high-stakes decision, making sure the high-stakes decision won't come as a surprise (Liebrechts 2022).

Step 4: assessors make high-stakes decisions based on multiple data points

To ensure that programmes are compliant with quality assurance processes, PA implies that credits will be awarded for the integrated block, rather than for single assessments or data points. High-stakes decisions are usually made annually or bi-annually (i.e. involving 30 or 60 credits). Sometimes programmes will choose a shorter or longer interval, such as quarterly or after 2 years of study. Degree apprenticeship programmes in the UK, which were introduced in 2015, have some similarity to PA by using formative assignments during the apprenticeship and

an end-point assessment at the end of the programme. However, unlike PA, assessment centre decisions usually focus on one-off events, rather than a review of evidence gathered longitudinally across the programme. That is, an assessment centre involves tasks that are additional to those undertaken during the programme, rather than an aggregation of tasks already completed for the degree award.

Robust decision-making (i.e. valid and reliable) is ensured by the large sample of data points, the involvement of multiple assessors, and multi-stage decision-making procedures. High-stakes decisions almost never come as a surprise. Concerns will already have been raised during medium-stakes decisions, in dialogue with the mentor. Multi-stage procedures could mean, for example, that a student's portfolio is first assessed by two independent assessors, who discuss their points of view, which builds moderation into the process. If they cannot come to agreement or express doubt about whether the student can pass, the student's portfolio is discussed in the entire assessment committee.

In the business programme presented above, students add a self-assessment to their portfolio, an evaluative judgement of their qualities and points for improvement, in relation to specified learning outcomes. The portfolio also contains a judgement by the student's mentor. For decision-making, an independent assessor (i.e. not the mentor) reviews the portfolio from a 'helicopter perspective': the student's overall development is judged, considering all data points in a holistic way. The assessor uses a holistic rubric based on the learning outcomes. The assessor does not re-assess all data points because many others have already given their feedback (i.e. their quality judgement) on these data points. Assessors thus look at patterns across the portfolio, considering whether this student – in general – is judged positively by important others. The independent assessor reaches a preliminary decision, after which the assessment committee, consisting of three members, discusses preliminary decisions for each student. When in doubt, they look at the data points to get a better impression of the student's work. The preliminary decisions are adapted when necessary. When the committee decides evidence is insufficient to meet standards on a given learning outcome, students may be permitted to progress but must collect more data points over a specified timeframe (e.g. one month) to demonstrate they have met the required standard. Thus, they will resubmit their portfolio including selected new evidence to the panel. In other cases, if students fall too far short of the standard on too many of the learning outcomes, students may need to repeat the term or year.

Implications of PA for HE leaders, managers, and professional services staff

Successful implementation of any educational innovation requires commitment and effort across HE leaders, managers, and a range of professional services staff working together, not just academic staff teaching on a programme. We address leadership required to support programmatic assessment, then discuss how professional services staff may be involved in the curricular design processes, and the ways other institutional systems may need to accommodate and support PA.

First, because PA conceptualises assessment at a programme, rather than module level, leadership across a programme and at an institutional level is vital. To achieve long-lasting and sustainable change, a double loop or renewal loop is needed (Argyris 2002). This means taking a step back to reflect on current practices and their associated goals, beliefs, and cultures. A shared and explicit vision on assessment and education is crucial. What do you – as a team – value as the core of learning, and how will PA allow you to support this learning better? On a practical level, leaders need to facilitate time and space. Experience shows that a few hours per week do not offer enough space for in-depth development and maturation.

Second, in the examples we have cited, programmes have often worked in curricular design teams consisting of academics teaching within the programme, as well as educational or assessment experts, administrators, and careers staff, who together have an overview of the entire programme and the professional work field. Educationalists and assessment experts offer perspectives from educational literature, and because they often work across programmes, they can share experiences from elsewhere. Careers and employability staff will be helpful collaborators in the design process because PA offers a unique opportunity to embed the development of students' career management competencies (Farenga and Quinlan 2016). PA focuses on developing and assessing competencies as demonstrated in a variety of career-relevant assessment activities. Career staff can ensure that the competences framework is aligned with employer requirements and help students identify and represent these career-related competences.

Third, in addition to professionalisation and guidance within the design process, policies and institution-wide cultures, regulations and systems can hinder or promote the paradigm shift towards PA. Writing institution-wide assessment and feedback strategies so that they promote or, at least, are consistent with the paradigm shift represented by PA is vital. Thus, Deputy and Pro Vice Chancellors for Education and their immediate managers of learning and teaching strategy who write policies and strategies need to be aware of and supportive of this new approach.

Quality assurance (QA) managers will need to be involved to ensure that changes made to programmes are aligned with regulations. For example, PA requires revising programme specifications to reflect the competences framework as well as combining and revising module guides and handbooks to reflect integrated content and skills described in relation to the competence framework and new assessment procedures. These changes trigger institutional review procedures in UK universities and affect contractual agreements with students that have timeline implications for rollout. QA staff and others who review changes in this key documentation need to be familiar with the principles of PA, especially related to the steps involved in making high stakes decisions that generate marks or progression. In some cases, QA staff may need to revise regulations to enable PA while assuring the university is still compliant with external regulations. Across a university, PA may be implemented in some programmes but not others, so the overall institutional framework, regulations, and procedures need to accommodate these differences. Some modification to descriptions of academic achievements on the UK's Higher Education Achievement Report (HEAR) will be needed to capture the competences framework and students' achievement on it.

Finally, educational technologists, learning development advisors and even student services staff who are all steeped in modular systems with traditional assessment systems may need to adjust their services to accommodate PA, ranging from integrating e-portfolio programmes into existing learning management systems to helping students extract competencies and prepare for medium-stakes dialogues to understanding new pressure-points in students' academic schedules. In short, any transformation in educational or assessment systems in higher education depends upon successful coordination of staff effort in a variety of roles across an institution.

Conclusion

PA is designed to promote learning, ensure rigour in assessment decision-making, and provide students with materials for and experience in communicating their employability. Thus, it offers a solution to three key challenges faced by the sector in assessment and feedback and, more generally, quality enhancement of learning and teaching. While innovative in the UK, it has been tested and refined in other contexts, providing examples of how it can be adapted to a variety of subjects. In short, we propose that PA can transform HE by creating a more productive assessment and feedback culture that is more fit for three key purposes: learning, assuring standards, and employability.

Note

1. Given our focus on student-centred practice, it is beyond the scope of this paper to discuss in detail how programmatic assessment may be used to address other specific purposes of assessment, such as institutional accreditation, professional licensing, qualifications frameworks or overall comparability of learning outcomes across institutions and countries. Furthermore, while we emphasise communicating employability as a key purpose of assessment, we recognise that higher education has broader purposes, and that assessment must be aligned to intended educational purposes. Programmatic assessment can be used to address whatever intended learning outcomes a programme values.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors



Liesbeth Baartman works as a professor of assessment in professional / vocational education at Utrecht University of Applied Sciences. The main focus of her (practice-based) research is on how assessment systems can capture the complexity of what students need in diverse workplaces. This includes research on programmatic assessment, workplace learning and assessment for learning.



Kathleen M. Quinlan is Professor of Higher Education and Director of the Centre for the Study of Higher Education at the University of Kent. Her research is broadly in the areas of teaching, learning, assessment and student engagement in higher education.

ORCID

Liesbeth K. J. Baartman  <http://orcid.org/0000-0002-3369-9992>

Kathleen M. Quinlan  <http://orcid.org/0000-0003-3606-4148>

References

- Argyris, C. 2002. "Double-Loop Learning, Teaching, and Research." *Academy of Management Learning and Education* 1 (2): 206–218.
- Baartman, L. K. J., H. Baukema, and F. J. Prins. 2022. "Exploring Students' Feedback Seeking Behavior in the Context of Programmatic Assessment." *Assessment & Evaluation in Higher Education* 48 (5): 598–612. <https://doi.org/10.1080/02602938.2022.2100875>.
- Baartman, L. K. J., T. van Schilt-Mol, and C. P. M. van der Vleuten. 2020. *Programmatisch toetsen. Voorbeelden en ervaringen uit de praktijk* [Programmatic Assessment. Examples and Practical Experiences]. Amsterdam: Boom.
- Baartman, L. K. J., T. van Schilt-Mol, and C. P. M. van der Vleuten. 2022. "Programmatic Assessment Design Choices in Nine Programs in Higher Education." *Frontiers in Education* 7:931980. <https://doi.org/10.3389/educ.2022.931980>.
- Biggs, J. 1996. "Enhancing Teaching through Constructive Alignment." *Higher Education* 32 (3): 347–364. <https://doi.org/10.1007/BF00138871>.
- Black, P., and D. Wiliam. 2018. "Classroom Assessment and Pedagogy." *Assessment in Education: Principles, Policy and Practice* 25 (6): 551–575. <https://doi.org/10.1080/0969594X.2018.1441807>.
- Boud, D., and E. Molloy. 2013. "Rethinking Models of Feedback for Learning: The Challenge of Design." *Assessment and Evaluation in Higher Education* 38 (6): 698–712.
- Carless, D. 2015. *Excellence in University Assessment. Learning from Award Winning Practice*. London: Routledge. <https://doi.org/10.4324/9781315740621>.
- Carless, D., and D. Boud. 2018. "The Development of Student Feedback Literacy: Enabling Uptake of Feedback." *Assessment & Evaluation in Higher Education* 43 (8): 1315–1325. <https://doi.org/10.1080/02602938.2018.1463354>.
- Constantinou, F. 2020. "What is Synoptic Assessment? Defining and Operationalising an as Yet Non-mainstream Assessment Concept." *Assessment in Education: Principles, Policy & Practice* 27 (6): 670–686. <https://doi.org/10.1080/0969594X.2020.1841734>.
- De Jong, L. H., H. G. J. Bok, W. D. J. Kremer, and C. P. M. van der Vleuten. 2019. "Programmatic Assessment: Can We Provide Evidence for Saturation of Information?" *Medical Teacher* 41 (6): 678–682. <https://doi.org/10.1080/0142159X.2018.1555369>.
- De Vos, M. E., L. K. J. Baartman, C. P. M. Van der Vleuten, and E. De Bruijn. 2022. "Unravelling Workplace Educators' Judgment Processes When Assessing Students' Performance at the Workplace." *Journal of Vocational Education & Training*, 1–20. <https://doi.org/10.1080/13636820.2022.2042722>.
- Driessen, E. W., J. van Tartwijk, M. J. B. Govaerts, P. Teunissen, and C. van der Vleuten. 2012. "The Use of Programmatic Assessment in the Clinical Workplace: A Maastricht Case Report." *Medical Education* 34:226–231. <https://doi.org/10.3109/0142159X.2012.652242>.
- Elkington, S. 2020. *Essential Frameworks for Enhancing Student Success: Transforming Assessment in Higher Education, A Guide to the Advance HE Framework*. York: Advance HE.
- Farenga, S. A., and K. M. Quinlan. 2016. "Classifying University Employability Strategies: Three Case Studies and Implications for Practice and Research." *Journal of Education and Work* 29 (7): 767–787. <https://doi.org/10.1080/13639080.2015.1064517>.
- Harland, T., A. McLean, R. Wass, E. Miller, and K. N. Sim. 2015. "An Assessment Arms Race and its Fallout: High Stakes Grading and the Case for Slow Scholarship." *Assessment and Evaluation in Higher Education* 40 (4): 528–541. <https://doi.org/10.1080/02602938.2014.931927>.
- Harrison, C. J., K. D. Könings, E. F. Dannefer, L. W. Schuwirth, V. Wass, and C. P. van der Vleuten. 2016. "Factors Influencing Students' Receptivity to Formative Feedback Emerging from Different Assessment Cultures." *Perspectives on Medical Education* 5:276–284. <https://doi.org/10.1007/S40037-016-0297-X>.
- Healy, M. 2023. "Careers and Employability Learning: Pedagogical Principles for Higher Education." *Studies in Higher Education*, Online first. <https://doi.org/10.1080/03075079.2023.2196997>.

- Heeneman, S., L. H. de Jong, L. J. Dawson, T. J. Wilkinson, A. Ryan, G. R. Tait, N. Rice, D. Torre, A. Freeman, and C. P. M. van der Vleuten. 2021. "Ottawa 2020 Consensus Statement for Programmatic Assessment – 1. Agreement on the Principles." *Medical Teacher* 43 (10): 1139–1148. <https://doi.org/10.1080/0142159X.2021.1957088>.
- Henderson, M., R. Ajjawi, D. Boud, and E. Molloy. 2019. "Identifying Feedback that has Impact." In *The Impact of Feedback in Higher Education: Improving Assessment Outcomes for Learners*, edited by M. in Henderson, R. Ajjawi, D. Boud, and E. Molloy, 15–34. Cham: Palgrave Macmillan.
- Higher Education Academy. 2015. "Higher Education Achievement Report". Accessed 26 May 2023. <https://www.hear.ac.uk/>.
- Horn, J. 2013. "Signature Pedagogy/Powerful Pedagogy: The Oxford Tutorial System in the Humanities." *Arts and Humanities in Higher Education* 12 (4), <https://doi.org/10.1177/1474022213483>.
- Imanipour, M., and M. Jalili. 2016. "Development of a Comprehensive Clinical Performance Assessment System for Nursing Students: A Programmatic Approach." *Japan Journal of Nursing Science* 13 (1): 46–54. <https://doi.org/10.1111/jjns.12085>.
- Jessop, T., and C. Tomas. 2017. "The Implications of Programme Assessment Patterns for Student Learning." *Assessment & Evaluation in Higher Education* 42 (6): 990–999. <https://doi.org/10.1080/02602938.2016.1217501>.
- Jorre de St Jorre, T., D. Boud, and E. D. Johnson. 2021. "Assessment for Distinctiveness: Recognising Diversity of Accomplishments." *Studies in Higher Education* 46 (7): 1371–1382. <https://doi.org/10.1080/03075079.2019.1689385>.
- Kickert, R., M. Meeuwisse, K. M. Stegers-Jager, P. Prinzie, and L. R. Arends. 2021. "Curricular Fit Perspective on Motivation in Higher Education." *Higher Education* 83:729–745. <https://doi.org/10.1007/s10734-021-00699-3>.
- Liebrechts, N. 2022. "Student and Teacher Feedback Literacy in the Context of Programmatic Assessment." Master's Thesis, Open University of the Netherlands. <https://research.ou.nl/en/studentTheses/student-and-teacher-feedback-literacy-in-the-context-of-programma>.
- Molloy, E., D. Boud, and M. Henderson. 2020. "Developing a Learning-centred Framework for Feedback Literacy." *Assessment & Evaluation in Higher Education* 45 (4): 527–540. <https://doi.org/10.1080/02602938.2019.1667955>.
- Nicol, D. 2021. "The Power of Internal Feedback: Exploiting Natural Comparison Processes." *Assessment and Evaluation in Higher Education* 46 (5): 756–778. <https://doi.org/10.1080/02602938.2020.1823314>.
- Oudkerk Pool, A., M. J. B. Govaerts, D. A. D. C. Jaarsma, and E. W. Driessen. 2018. "From Aggregation to Interpretation: How Assessors Judge Complex Data in a Competency-based Portfolio." *Advances in Health Sciences Education* 23 (2): 275–287. <https://doi.org/10.1007/s10459-017-9793-y>.
- Pitt, E., and K. M. Quinlan. 2022. *Impacts of Higher Education Assessment and Feedback Policy and Practice on Students: A Review of the Literature 2016-2021*. York, UK: Advance HE. <https://www.advance-he.ac.uk/knowledge-hub/impacts-higher-education-assessment-and-feedback-policy-and-practice-students-review>.
- Sadler, D. R. 1989. "Formative Assessment and the Design of Instructional Systems." *Instructional Science* 18 (2): 119–144. <https://doi.org/10.1007/BF00117714>.
- Sadler, D. R. 2007. "Perils in the Meticulous Specification of Goals and Assessment Criteria." *Assessment in Education: Principles, Policy and Practice* 14 (3): 387–392. <https://doi.org/10.1080/09695940701592097>.
- Schut, S., L. A. Maggio, S. Heeneman, J. van Tartwijk, C. van der Vleuten, and E. Driessen. 2021. "Where the Rubber Meets the Road—An Integrative Review of Programmatic Assessment in Health Care Professions Education." *Perspectives in Medical Education* 10:6–13. <https://doi.org/10.1007/S40037-020-00625-W>.
- Stobart, G. 2008. *Testing Times: The Uses and Abuses of Assessment*. London: Routledge. <https://doi.org/10.4324/9780203930502>.
- Torre, D., N. E. Rice, A. Ryan, H. Bok, L. J. Dawson, B. Bierer, et al. 2021. "Ottawa 2020 Consensus Statements for Programmatic Assessment – 2. Implementation and Practice." *Medical Teacher* 43 (10): 1149–1160. <https://doi.org/10.1080/0142159X.2021.1956681>.
- Torre, D. M., L. W. T. Schuwirth, and C. P. M. van der Vleuten. 2020. "Theoretical Considerations on Programmatic Assessment." *Medical Teacher* 42 (2). <https://doi.org/10.1080/0142159X.2019.1672863>.
- Van der Vleuten, C. P. M., L. W. T. Schuwirth, E. W. Driessen, J. Dijkstra, D. Tigelaar, L. K. J. Baartman, et al. 2012. "A Model for Programmatic Assessment fit for Purpose." *Medical Teacher* 34 (3): 205–214. <https://doi.org/10.3109/0142159X.2012.652239>.
- Van der Vleuten, C. P. M., L. T. W. Schuwirth, F. Scheele, E. W. Driessen, and B. Hodges. 2010. "The Assessment of Professional Competence: Building Blocks for Theory Development." *Best Pract. Res. Clin. Obstet. Gynaecol* 24 (6): 703–719. <https://doi.org/10.1016/j.bpobgyn.2010.04.001>.
- Wilkinson, T. J., M. J. Tweed, T. G. Egan, A. N. Ali, J. M. McKenzie, M. Moore, et al. 2011. "Joining the Dots: Conditional Pass and Programmatic Assessment Enhances Recognition of Problems with Professionalism and Factors Hampering Student Progress." *BMC Medical Education* 11 (1): 29. <https://doi.org/10.1186/1472-6920-11-29>.
- Winstone, N., and D. Boud. 2019. "Exploring Cultures of Feedback Practice: The Adoption of Learning Focused Feedback Practices in the UK and Australia." *Higher Education Research and Development* 38 (2): 411–425. <https://doi.org/10.1080/07294360.2018.1532985>.