



Managing the Shift to Programme-Level Assessment: Structural Barriers, Stakeholder Experiences, and Mechanisms of Sustained Transition

Authors: Nigel Francis, Pam Birtill, Mark Davies, Peter Hartley, Ryan James, Stephen Rutherford, David Tree

This Literature Review is an output from a [Collaborative Enhancement Project](#) supported and funded by the QAA.

The project was led by University of South Wales, in partnership with Brunel University, Cardiff University, Edge Hill University and University of Leeds. More information about the project can be found [here](#).

Abstract

Programme-level assessment (PLA) has gained traction in higher education policy and practice, yet implementation consistently falls short of stated ambitions. This systematic review synthesises 25 empirical and conceptual studies to examine the structural and governance factors shaping PLA implementation across institutional and disciplinary contexts. We find that implementation failure is primarily architectural: governance systems, quality assurance frameworks, and credit structures were designed to reproduce modular logic, and staff resistance, though real, is downstream of this more fundamental problem. The two principal models of PLA, programmatic assessment, developed in medical and health professional education, and integrated programme assessment (IPA), emerging across broader disciplines, are analytically distinct rather than interchangeable approaches to a shared problem. Both face the same institutional architectural constraints but with structurally different resources to address it. The mechanisms that sustain transition work by redistributing the cost–benefit asymmetry of implementation - the upfront costs that are faced by different people than those that feel the downstream benefits - rather than through efforts to change attitudes or secure individual buy-in. The evidence base remains heavily concentrated in health and professional education, and findings should be applied to other disciplinary contexts with caution.

Introduction

The imperative for change: addressing the fragmented curriculum

Assessment is widely acknowledged as a primary driver of student behaviour in higher education: it defines the curriculum for students, dictating what they regard as necessary and how they allocate their time and effort (Boud, 2000; Rust, 2002), and if institutions wish to improve the quality of student learning they must fundamentally alter how they assess it (Brown et al., 2013). Yet traditional assessment practices remain essentially modular and fragmented. The dominant structure of university education, organised into discrete, self-contained modules, has cultivated a disjointed learning experience (Jessop & Tomas, 2017) in which students are encouraged to revise and forget and fail to see the overarching rationale of their programme of study (Hartley & Whitfield, 2012). Students perceive assessments not as steps in a developmental journey but as isolated hurdles to be cleared and discarded (Eva et al., 2016), incentivising surface approaches to learning and the compartmentalisation of knowledge rather than the synthesis of ideas and holistic competencies required in professional environments (Boud & Molloy, 2013).

Short, compressed semesters and high volumes of summative assessment create an assessment arms race in which modules compete for student attention at the expense of deep engagement (Harland et al., 2015). When feedback is confined to a single module, students struggle to feed forward insights from one task to the next, limiting opportunities for the cumulative development of self-regulated learning and evaluative judgement across a programme (Rutherford, 2019; Hughes et al., 2015). To address these systemic failures, there is a growing interest in Programme-Level Assessment (PLA).

Defining the shift: from modular to holistic assessment

In higher education practice, two main approaches to PLA can be distinguished: programmatic assessment and integrated programme assessment (IPA, see figure 1). Programmatic assessment was developed and is largely confined to medical education, in which assessment has been conceptualised as a system of a large number of longitudinally connected data points rather than a series of disconnected tests (van der Vleuten et al., 2012, figure 1D). In programmatic assessment models, individual assessment points are optimised for learning (low stakes), while high-stakes progression decisions are based on aggregated information across the entire programme (Schuwirth & Van der Vleuten, 2011). IPA draws on similar concerns for coherence, longitudinal development and meaningful progression decisions, but has emerged within a broader subject base with an emphasis on aligning fewer, higher-stakes assessments which integrate knowledge from across taught modules to support and assess cumulative learning and programme outcomes (figure 1 B, C).

PLA involves strategically designing assessments to sequentially develop over-arching discipline knowledge, transferable skills, and career readiness (Charlton & Newsham-West, 2024a). This approach aligns with the concept of constructive alignment, ensuring that teaching activities and assessment tasks are explicitly linked not only to useful module objectives but also to high-level programme outcomes (Biggs, 1996). By integrating assessment strategies, such as longitudinal portfolios and synoptic or integrative assessments, PLA aims to foster connective learning (Hartley & Whitfield, 2012). Such assessments facilitate students to develop evidence of their learning over time, facilitating feedback processes that bridge the gap between formative and summative assessment (Clarke & Boud, 2018). In this sense, IPA can be seen as running in parallel with programmatic assessment: both respond to concerns about fragmentation and one-off testing, but they do so through distinct design practices and with different degrees of emphasis on formalised, portfolio-based decision-making.

Where PLA has been successfully implemented, the benefits are well-documented across stakeholder groups. For students, PLA mitigates the fragmented learn-and-

forget cycles encouraged by modular structures, developing feedback literacy, self-regulation, and evaluative judgement over time, transitioning students from passive recipients of grades to proactive, lifelong learners (Baartman et al., 2022; Baartman & Quinlan, 2024). The reduction in overall assessment volume that typically accompanies PLA has been shown to improve student psychological wellbeing and lower assessment-related anxiety (Vanderheide & Walkington, 2009; Ryan & Judd, 2022). For academic staff, well-implemented PLA reduces summative marking loads and has been shown to break down entrenched disciplinary silos, encouraging collaboration and a shared sense of curricular ownership (Schwegler et al., 2024). For employers and external stakeholders, PLA provides a more authentic and holistic representation of graduate competencies than module-by-module transcripts can offer (Kotnour et al., 2022; Tree et al., 2025). In this sense PLA embodies the principle of sustainable assessment, equipping students not merely to pass their degree but to manage their own learning beyond it (Boud, 2000; Boud & Soler, 2016). The question this review addresses is not whether PLA is worth pursuing as the evidence suggests it is, but why implementation so consistently falls short of these ambitions.

The implementation challenge: culture, policy, and stakeholders

While the pedagogical arguments for PLA are compelling, managing the shift from modular to programme-level assessment presents significant operational and cultural challenges. Higher education institutions (HEIs) are characterised by many rigid infrastructures: student record systems, timetabling, and academic regulations are frequently designed for modular granularity rather than programmatic integration (Charlton & Newsham-West, 2024b). For example, university policies regarding grade aggregation and credit accumulation often embed the siloed, modular approach, making it challenging to implement synoptic assessments that span multiple units of study (Jessop et al., 2012). These structural features of institutional architecture represent the primary obstacle to PLA implementation, a point this review develops in detail in the findings.

The transition to PLA necessitates a significant cultural transformation among key stakeholders. Academic staff, accustomed to autonomy within their own modules, may view programme-level planning as a threat to their academic freedom or as an imposition of additional workload (Jessop et al., 2012). Recent studies identify isolated practices, resource constraints, and a lack of assessment literacy among staff as significant barriers to PLA adoption. While these cultural and attitudinal factors are real and well-documented, this review argues that they are downstream of the structural problem: staff resistance is in significant part a rational response to being asked to implement change within a system that has not been restructured to enable it.

Staff must move from being the sole arbiters of a specific subject to collaborators in a shared educational environment, necessitating a shift that demands effective leadership and professional development (Charlton & Newsham-West, 2024b).

For students, the shift is equally impactful. While PLA promises a more coherent learning experience, it also introduces new anxieties regarding the perceived high-stakes nature of integrative assessments. Students often prefer multiple, smaller assessments to mitigate risk, even though they recognise that such patterns lead to fragmented learning (Pye et al., 2025). The evidence reviewed here suggests, however, that this initial anxiety is transitional rather than fundamental: where assessment architecture is stable and well-supported, anxiety gives way to improved engagement, resilience, and outcomes. Therefore, implementing PLA is not merely a technical redesign but a complex change-management process that involves negotiating diverse stakeholder interests.

Rationale for the review

Despite a growing body of literature advocating programme-focused assessment, there remains a gap in understanding the practical mechanics of implementation beyond medical education, which tends to focus more on programmatic assessment. While the theoretical benefits are well documented (van der Vleuten et al., 2012), there is a scarcity of systematic analyses of how institutions manage barriers related to policy, resources, and stakeholder resistance across broader disciplines. Specifically, it is necessary to understand how the tensions between administrative efficiency and pedagogical effectiveness are navigated during this transition (Charlton & Newsham-West, 2024b).

This systematic review addresses this gap by moving beyond the pedagogical 'why' to explore the operational 'how.' It does so by advancing three arguments that the existing literature has not stated explicitly: that implementation failure is primarily architectural rather than cultural; that programmatic assessment and IPA are analytically distinct models with structurally different capacities to address the architecture problem; and that the mechanisms which sustain transition work by redistributing the cost-benefit asymmetry of implementation rather than by asking individuals to absorb it. By synthesising evidence on regulatory barriers, workload implications, and stakeholder experiences, this review aims not only to map the implementation landscape but to reframe the question institutions and policymakers need to ask.

Research questions

To guide this inquiry, this review addresses the following research questions:

1. What are the primary structural and cultural barriers to implementing programme-level assessment in higher education, and how does the relationship between them shape implementation outcomes?
2. How does the shift to programme-level assessment impact the key stakeholders (students, academic staff, and administrators) in terms of workload, engagement, and satisfaction?
3. What effective management strategies, governance structures, and policy changes have been identified to facilitate the successful transition to sustainable programme-level assessment?

Methods

Comprehensive literature searching

Systematic searches across two databases, Scopus and ProQuest, were carried out using Boolean search strings (Table 1). Each string was matched as closely as possible to each other within the constraints of the requirements of the individual databases, to ensure alignment between searches. The search strings were designed to identify papers that evaluated programme-level assessment in higher education, with a specific focus on adoption and implementation processes and stakeholder experiences.

Further literature was identified through backwards and forwards snowballing. Backwards snowballing was achieved by scanning reference lists of papers to identify older, foundational papers. To facilitate the forward snowballing process, where citations of seed papers are identified, the AI-powered tool Research Rabbit was used. In addition to papers identified through database searches and snowballing some work, including two sector reports, was included on the basis of direct relevance identified during the review process.

Systematic refinement using inclusion and exclusion criteria

In total, 663 papers were identified through Scopus and ProQuest searches (last date of screening: 16th October 2025). Each paper was subsequently reviewed for relevance to the topic, initially by title only and subsequently by abstract. From these screened papers, further papers were identified through snowballing. Following these two filtering steps, exclusion criteria (Table 2) were applied to the remaining papers in accordance with the PRISMA analysis protocol (Page et al., 2021).

Numbers of records excluded and retained at each stage are summarised in a PRISMA diagram (Figure 2). Primary screening was conducted using the filter features of the databases themselves to exclude papers that were not primary research papers, not available in English, or not available to access. Each remaining paper was evaluated independently by the same three members of the research team against the exclusion criteria in Table 2. Initially, the title and abstract were reviewed, and then any remaining papers were read in full and evaluated. Where differences of opinion occurred, a fourth member of the research team reviewed the paper independently to provide the final decision. After the application of exclusion criteria, the final number of papers included in the review was 25.

A thematic analysis process was adopted, which was adapted from the 6-stage structure described by Naeem et al. (2023). Step 1: Transcription and familiarisation with the data was undertaken via detailed reading of the papers. Step 2: Keywords identified linking to the implementation of IPA. Step 3: Coding undertaken through an in-depth review of each paper, noting specific examples. Step 4: Common themes identified. Step 5, keywords, codes, and themes compiled and conceptualised. Step 6: Development of the conceptual model.

Findings

The findings are organised around the three research questions (see Table 3 for a summary of all included studies). The first section examines the structural and governance barriers to PLA implementation, arguing that institutional architecture rather than staff attitudes is the primary obstacle. The second section draws together evidence on stakeholder experience, focusing on the cost-benefit asymmetry that shapes how staff and students experience the transition. The third section examines the management strategies and governance mechanisms that have been found to sustain transition, arguing that what works does so by redistributing implementation costs.

Research Question 1: Barriers to implementation

1.1 The institutional architecture of higher education reproduces modular logic

The widespread adoption of modular degree structures represents one of the most significant structural barriers to PLA. Jessop (2023) and Jimenez et al. (2022) both discuss how modularisation inherently compresses learning into short timeframes, such as 10-15-week terms or semesters, thereby forcing a fragmented, "atomised" approach to assessment, in which modules are often too short to facilitate complex, deep learning.

Fielding and Regehr (2017) describe how this structural fragmentation encourages a memorise-and-forget learning cycle, in which students rote-learn content for isolated, end-of-module exams, only to quickly forget it once the module concludes. Kovaichelvan and King (2020) suggests that traditional structures encourage this rote-learning and fail to support higher-order synthesis. Consequently, Jessop (2023) highlights that both staff and students often fail to recognise the overarching coherence and aims of the educational programme, viewing assessment merely as a disconnected series of hurdles.

These structural features of modular assessment are well recognised and yet we will argue that the literature has misdiagnosed the primary obstacle to PLA implementation. The existing literature consistently identifies a cluster of obstacles to PLA: staff resistance, student anxiety, resource constraints, and inadequate leadership. While these exist we argue that treating them as the primary explanation for implementation failure is a misdiagnosis. What the literature shows, when examined across contexts and model types, is something more structurally intractable: institutions attempting to implement PLA are doing so using governance systems that were specifically designed to reproduce the modular logic they are trying to escape.

Resistance by staff is therefore a consequence of fundamental misfit between institutional architecture and the demands to PLA, rather than being intrinsically against change.

Quality assurance and restrictive policy frameworks

The clearest evidence for this argument comes from the role of institutional quality assurance frameworks and validation processes. Jessop & Gubby (2012) points to the imposition of rigid, regulations that promote a one-size-fits-all approach. QA frameworks were also a focus for Gibbs and Dunbar-Goddet (2009), who discussed how they tend to focus heavily on standards, metrics, and certification rather than the pedagogical mechanisms that actively support student learning. Jessop et al. (2012) also highlighted how programme and module validation documents are heavily skewed towards summative assessments, rendering formative, developmental assessment as less effectively specified and, by implication, discretionary.

Bowe and Armstrong (2017) describe an embedded culture of accreditation, which places excessive reliance on the outcomes of segregated modules, leading to assessment fatigue and failing to capture holistic student learning and progression across modules. This finding is compounded by university policies, where assessments are frequently viewed exclusively at the individual module level and lack the explicit terminology and progressive guidelines needed to mandate cross-module integration of synoptic assessment (Charlton et al., 2022; Charlton and Newsham-West, 2024b).

Gibbs and Dunbar-Goddet (2009) extended this point through a detailed case study of nine undergraduate degree programmes across three contrasting UK universities, and their findings show that QA frameworks direct institutional attention toward standards, alignment metrics, and summative compliance rather than toward the features of assessment that are known to support student learning, such as formative assessment and timely feedback. In doing so, they actively orient institutional energy away from the programme-level coherence that PLA requires and toward a form of module-by-module compliance that leaves the architecture of learning largely unexamined.

Assessment policy language and the reproduction of modularity

The problem is compounded by the language of assessment policy itself. Charlton et al. (2022) found that assessment policies routinely lacked the explicit, consistent terminology needed to mandate cross-module integration. Terms such as 'holistic', 'progressive', and 'vertical alignment' appeared inconsistently or not at all, meaning that policies which nominally endorsed programme-level approaches provided no operational mechanism for actioning them. Assessment continued to be evaluated and governed at the level of individual modules by default, because the policy architecture contained no other unit of analysis.

In Australia, TEQSA's Higher Education Standards Framework requires that course design enable the "progressive and coherent achievement" of course-level learning outcomes (Standard 3.1.3; TEQSA, 2017). This requirement is broadly compatible with PLA principles, but it can be satisfied through documentation of existing modular assessment structures without requiring genuine programme-level integration, creating the appearance of alignment while leaving the underlying architecture of modular assessment intact.

Locally negotiated workarounds and their fragility

An illustration of how institutional architecture forces practitioners to work around rather than through existing governance systems comes from Tree et al. (2025). In two of the four case study institutions examined, BA Journalism at Sheffield University and the Foundation Year in Nursing at the University of Surrey, programme teams created single 120-credit modules for an entire year of study. This allowed integrated teaching and assessment to proceed while technically remaining compliant with the institutions' modular systems. This remains a pragmatic workaround rather than a structural solution: locally negotiated arrangements of this kind are dependent on the goodwill of professional services staff and the continuity of personal relationships that may be disrupted when key individuals move on.

A university-wide solution, understood by both academic and professional staff, is identified as preferable, but in most contexts, no such solution yet exists. The creation of 120-credit synthetic modules to circumvent modular governance is not evidence of successful PLA implementation; it is evidence that the institutional architecture is sufficiently hostile to PLA that practitioners must work around it rather than through it. The workaround sustains the change at programme level while leaving the underlying architectural problem entirely intact. The fragility of such arrangements is further evidenced by the finding that practitioners across seven Australian universities identified three times as many barriers as enablers to programme-level assessment, with the barriers being structural and the enablers almost entirely personal and relational (Charlton & Newsham-West, 2024b). Even the most developed available model of institutional support, the Nottingham Civil Engineering redesign, required dedicated resources, external accreditation leverage, and a willingness of staff to take on significant extra workload in the first iteration before the structural benefits materialised (Turnbull, 2020).

Staff resistance as a structural symptom

The literature indicates that it would be wrong to suggest that governance architecture is the only obstacle, or that staff attitudes are irrelevant. The cultural dimensions of this transition, including the cost-benefit asymmetry that shapes how academic staff and students experience change over time, are real and significant. They are examined in detail in Section 3. What the architectural evidence establishes is that even where staff are willing to engage with PLA, they routinely encounter institutional systems that make it structurally difficult to do so. The resistance that appears cultural is often, in significant part, a rational response to being asked to implement change within a system that has not been restructured to enable it (Schwegler et al., 2024). This pattern was documented directly in the experiences of practitioners who found the forces obstructing implementation to be systemic and outside their control, while the forces enabling it depended entirely on personal relationships and individual goodwill (Charlton & Newsham-West, 2024b).

1.2 Programmatic and Integrated Programme Assessment: Two models of PLA with structurally different capacities

The significance of the institutional architecture problem becomes clearest when the two principal models of PLA are examined alongside each other. Programmatic assessment, developed within medical education (van der Vleuten et al., 2012; Schuwirth & Van der Vleuten, 2011), achieved its earliest and most sustained traction in a context where the institutional governance infrastructure was already, in important

respects, compatible with longitudinal, integrated assessment. Medical schools often operate within fitness-to-practise frameworks, manage progression through committee-based decisions including the institutional separation of the mentor role from the assessor role (Fielding & Regehr, 2017), and are accustomed to portfolio-based licensing requirements which all represent a pre-existing institutional architecture that is considerably less hostile to programme-level thinking than the credit-based modular frameworks governing most of broader higher education. IPA in broader HE disciplines faces a more fundamental architectural mismatch, which may explain why implementation evidence from those contexts shows more friction and slower adoption (Tree et al., 2025; Charlton & Newsham-West, 2024b; Charlton et al., 2022).

The literature frequently treats these two models as points on a continuum, related approaches to a shared problem, differing in degree rather than in kind. We argue that this framing obscures more than it illuminates. When the implementation evidence is examined, programmatic assessment and IPA reveal not just different design emphases but meaningfully different institutional conditions, different governance requirements, and a different capacity to address the architecture problem structurally. The distinction is not a terminological preference but an analytical necessity.

1.2.1 Programmatic assessment

A coherent framework with institutional roots

Programmatic assessment has a relatively clearly defined theoretical identity. Developed by van der Vleuten, Schuwirth and colleagues from the early 2000s, and given its most authoritative articulation in van der Vleuten et al. (2012), it rests on a specific conceptual logic: individual assessment data points are optimised for learning rather than judgment (low stakes), while high-stakes progression decisions are made by independent expert committees on the basis of aggregated portfolio evidence across the whole programme. The Ottawa 2020 Consensus Statement subsequently codified this logic into twelve agreed principles, providing the field with a shared framework of a kind that few educational innovations achieve (Heeneman et al., 2021). This degree of theoretical consolidation matters for implementation: it gives institutions a coherent blueprint to work from, and it gives practitioners a shared language for discussing and defending their design decisions.

The evidence base for programmatic assessment is correspondingly more developed than for IPA, though it is concentrated heavily in medical, pharmacy, and clinical education. Ryan and Judd (2022) document a staged transition across three phases in a medical education context, providing longitudinal evidence of what implementation actually looks like over time.

Medina et al. (2013) describe the incremental introduction of integrated assessment methods into a pharmacy curriculum over several years. Baartman et al. (2022) offer the most disciplinarily diverse evidence within the programmatic assessment literature, spanning health sciences, paramedic, teacher education, and communication. All nine programmes adhered to core principles of programmatic assessment including a backbone of learning outcomes, longitudinally connected data points, no pass/fail decisions on individual assessments, and committee-based high-stakes decision-making, but also made different design choices depending on professional context and practical constraints. The study also documents persistent difficulties such as students experiencing low-stakes data points as high-stakes, teachers resisting relinquishing module-level ownership; and in one case, an inability to agree on a shared backbone of learning outcomes. This confirms that programmatic assessment can extend beyond medical education but also that the institutional and cultural conditions enabling it are not necessarily present even in contexts like the Netherlands .

The governance infrastructure that makes it work and who can build it

The implementation profile that emerges from these papers is distinctive. Programmatic assessment does not simply redesign individual assessment tasks; it requires the construction of parallel governance infrastructure. Expert progression committees must be established, convened, and maintained. Portfolio systems must be built and integrated with student record systems. The separation of the mentor role from the judge role, identified by both Fielding and Regehr (2017) and van der Vleuten et al. (2012) as essential to the integrity of programmatic assessment, must be operationalised through institutional structures that most universities do not have by default. These represent a significant institutional investment, one that the evidence from medical and clinical education suggests can be achieved (Ryan & Judd, 2022; Baartman et al., 2022; van der Vleuten et al., 2012).

The evidence from medical education suggests that this investment has been made and the question of why it could be made in that context is important. Medical schools in research-intensive universities occupy an institutional position that is, in many respects, distinctive. They typically manage substantial budgets, maintain direct regulatory relationships with professional bodies and clinical licensing authorities that sit largely outside normal university governance, and carry sufficient reputational and financial weight to negotiate arrangements that other faculties cannot. This reflects a structural reality in which medical schools have the institutional standing to build parallel governance infrastructure without requiring the central university architecture to change around them.

This observation does not imply that programmatic assessment is not possible outside medicine. Baartman et al.'s (2022) study of nine Dutch programmes is significant precisely because it demonstrates that the model can travel beyond its medical education origins. The Dutch non-medical programmes examined by Baartman et al. (2022) are therefore not counterexamples to the institutional capacity argument but illustrations of what becomes possible in a policy environment where professional higher education institutions have historically had greater curricular autonomy and stronger traditions of competency-based qualification frameworks than the UK or Australian systems that dominate much of the IPA evidence base. The lesson is not that programmatic assessment is uniquely medical, but that its implementation requires a degree of institutional capacity and governance autonomy that has been most consistently available in medical and clinical education, and in a small number of other professional and disciplinary contexts with analogous structural features.

Residual tensions even where implementation has succeeded

The evidence for programmatic assessment is not uniformly positive and even in contexts where implementation has been sustained, the evidence reveals persistent tensions. Ryan and Judd (2022) discuss the difficulty of the transition: the logistical burden of managing longitudinal data and the challenge of maintaining the integrity of the low-stakes/high-stakes distinction under institutional pressure to use assessment data for accountability rather than learning. Ryan and Judd are explicit that full programmatic assessment, specifically, decision-making about progression based on holistic judgement of a student's competencies across attributes rather than aggregation of marks from individual assessments, "would not be permissible within our current university-wide assessment policy." Ryan and Judd's programme illustrates the architecture problem : central university policy remains a structural ceiling even for the most advanced implementations by a dedicated team. Fielding and Regehr (2017) identify a related structural tension that formal governance arrangements mitigate but do not eliminate: the challenge of moving from a model in which assessment is the exclusive responsibility and right of individual course directors to one in which assessment is a collective responsibility of the programme. Taken together, the evidence from programmatic assessment is not a story of solved problems but of challenges partially addressed in contexts where institutional investment was structurally possible, with significant structural tensions remaining even where that investment has been most extensive.

1.2.2 IPA: bottom-up origins, absent policy mandate, and disciplinary concentration

Bottom-up origins and the absence of a shared framework

The IPA evidence base has a markedly different character. Where programmatic assessment developed from a relatively coherent theoretical project codified in a widely cited consensus framework, IPA emerged more organically from practitioner-led sector development. Its intellectual foundations lie substantially in the Programme Assessment Strategies (PASS) project, a three-year NTFS-funded consortium led by Bradford, whose final report (Hartley & Whitfield, 2012) established the conceptual and practical basis for programme-focussed assessment across disciplines. Jessop's TESTA methodology (Jessop, 2023) represents a parallel and complementary strand, providing evidence-based programme audit tools that have been widely used to surface assessment patterns and stimulate programme-level conversations, though its primary contribution is diagnostic rather than a direct antecedent to IPA design. From these roots, IPA has spread through workshops, sector networks, and discipline-level initiatives rather than through a shared theoretical mandate. Tree et al. (2025) describe initiatives beginning in Biomedical Sciences at Brunel, spreading gradually to disciplines ranging from journalism to civil engineering to nursing. Charlton and Newsham-West (2024a, 2024b) and Charlton et al. (2022) document IPA-adjacent practices in Australian higher education, again emerging from discipline-level initiatives rather than institutional mandates.

This bottom-up character is both a strength and a vulnerability. The strength is adaptability: IPA has shown itself capable of operating across a wide range of disciplinary and institutional contexts, precisely because it is not tied to a single design template. Programmes in health promotion (Dukhi et al., 2014), sociology (Crockett et al., 2018), nursing (Vanderheide & Walkington, 2009), and databases computing (Gorra et al., 2007) have all developed IPA-compatible approaches that reflect the specific pedagogical and professional requirements of their disciplines. An integrative engineering programme developed at a US liberal arts college required fifteen years of governance negotiation and survived a failed faculty vote driven by resource concerns and institutional politics. Success came only through sustained relationship-building and transparent process, but the programme remained dependent on staffing arrangements that had not been fully resolved at the point of publication (Brandes & Anderson, 2020). This flexibility distinguishes IPA from programmatic assessment's more formalised and prescriptive framework. The vulnerability remains though and bottom-up approaches seem to reach a structural ceiling where sustained progress requires institutional mandate.

The vulnerability of IPA to institutional systems stems partly from the absence of an IPA equivalent to the Ottawa Consensus means there is no shared framework to anchor institutional conversations, no agreed set of principles against which implementation quality can be evaluated, and no common language for practitioners across disciplines to communicate what they are doing and why. The scale of this policy vacuum has been documented empirically. Charlton et al. (2022) conducted a systematic discourse analysis of 61 assessment-related policies from 22 top-ranked Australian universities, among the most likely institutions in the sector to be implementing best practice and found that only two explicitly used the terms 'program-based' or 'program-approach' to assessment. The term most clearly associated with programme-level thinking, vertical and horizontal alignment of assessments across a degree, appeared in the policies of just one institution. The dominant policy language was alignment, deployed overwhelmingly at module level rather than programme level. This means that in most institutions where IPA practitioners are attempting to implement programme-level approaches, the institutional policies that govern their work do not recognise, define, or require what they are doing. IPA practitioners must make the case from first principles each time, in each institution, often against governance frameworks that have no category for what they are proposing. Whereas programmatic assessment practitioners can invoke the Ottawa Consensus to legitimise their design decisions, IPA practitioners are working in a policy environment that is, in most institutions, effectively silent on the matter.

Dependence on goodwill rather than mandate

The implementation profile that emerges from the IPA evidence is characterised by a fundamental structural vulnerability: dependence on locally negotiated arrangements rather than institutionally mandated frameworks. The pragmatic solutions found such as creating 120-credit modules, negotiating exceptions with professional services, building informal relationships with registry staff, work within specific institutional relationships are fragile when those relationships change (Schwegler et al., 2024; Tree et al., 2025). The most direct empirical measure of this fragility comes from a study of 18 Associate Deans and Programme Directors across seven Australian universities, in which practitioners identified three times as many barriers to programme-level assessment as enablers (Charlton & Newsham-West, 2024b). Critically, the barriers were structural, restrictive policies, resource constraints, and ownership disputes rooted in modular governance, while the enablers were all relational and personal. The things that make IPA work were personal; the things that stopped it working are systemic. This fragility is compounded by workforce casualisation.

As institutions rely more heavily on sessional and contract staff, the sustained collaboration and shared programme vision that IPA requires becomes structurally harder to maintain, a pattern documented in both Australian and US contexts (Charlton & Newsham-West, 2024b; Schwegler et al., 2024).

A thinner and more discipline-concentrated evidence base

The IPA evidence base is disciplinarily narrow. The studies reviewed here are drawn predominantly from nursing and health-related disciplines, with smaller representation from social sciences, computing, and engineering. Sector-level guidance documents and institutional strategic initiatives confirm that programme-level assessment is being actively advocated and implemented across a broader disciplinary range, including humanities, social sciences, and arts (Walker, 2025), yet these disciplines are almost entirely absent from the peer-reviewed literature. The disciplinary concentration of the evidence base is acknowledged even within it: Charlton and Newsham-West's (2024a) conceptual model for programme-level assessment is the most developed framework of its kind and was built from participants in health and science disciplines exclusively, and its authors identify testing across a wider range of disciplines as a task future research will need to undertake. A conceptual model whose development excluded the disciplines where IPA is now most actively promoted cannot yet serve as a reliable guide for implementation in those contexts. The disciplinary reach of current IPA advocacy and the disciplinary scope of the evidence base do not yet match.

What the comparison reveals

Placed alongside each other, the two models face the same underlying problem, governance systems designed for modularity that must be circumvented or overridden to implement programme-level approaches, but with structurally different resources to address it. Programmatic assessment has largely solved this by building parallel governance infrastructure in contexts where medical schools possessed the institutional standing to do so. IPA in broader HE cannot easily do the same and relies instead on locally negotiated workarounds that are pragmatic but structurally fragile. Advice focused primarily on change management and incremental rollout addresses the symptoms of this problem rather than its cause. How institutions have begun to address the architecture/structural problem, and what mechanisms have proved most effective, is the subject of the following section.

RQ 2. Stakeholder experiences of transition

The evidence on stakeholder experience runs through both the structural barriers documented above and the transition mechanisms examined in the section that follows. This section draws that evidence together. The cost-benefit asymmetry that shapes how academic staff experience PLA implementation, and the anxiety-to-engagement trajectory that characterises student experience are central to understanding why implementation fails or succeeds.

Staff experience and the cost-benefit asymmetry

The structural barriers above do not operate in isolation from human experience. Staff and student resistance is real and well-documented, but as established in Section 1 it is best understood as a rational response to the structural problem rather than an independent obstacle. In this section we draw on the evidence on how that experience unfolds. Charlton and Newsham-West's (2024b) study captures the lived experience of this asymmetry directly: practitioners reported that the forces obstructing implementation were systemic and outside their control, while the forces enabling it depended entirely on personal relationships and individual goodwill. This describes a specific professional experience in which the capacity to act is structurally constrained while the responsibility to act remains personal. The temporal dimension is also significant: staff workload increased substantially in the first iteration of one documented redesign, while the benefits of reduced exam volumes, stronger student outcomes, more meaningful relationships with students all arrived later and accrued differently across the programme team (Turnbull, 2020). The costs of transition are real, front-loaded, and unevenly distributed while the benefits are real but delayed. That structural fact, more than any cultural disposition, seems to shape how staff experience and respond to PLA implementation.

Student experience: from anxiety to engagement

Student experience of PLA transition follows a consistent pattern across both model types. In the initial phase, students accustomed to modular, compartmentalised assessment encounter longitudinal, integrative demands for which their prior experience has not prepared them. Initial stress is well-documented: nursing students have reported acute anxiety when integrated assessments linked multiple modules, fearing that a single shared task failure would damage marks across several areas (Vanderheide & Walkington, 2009); even low-stakes formative tasks in programmatic contexts initially carry high-stakes feelings for students preconditioned by modular grading cultures (Baartman et al., 2022); and in one documented redesign the first

cohort was probably asked for too much (Turnbull, 2020). What the evidence also consistently shows is that the initial anxiety gives way when the assessment architecture is stable and well-supported to give significantly improved engagement, confidence, and outcomes. Students shift from stress to engagement as they become accustomed to PLA, with increased resilience and satisfaction over time (Vanderheide & Walkington, 2009), and measurably better outcomes follow in later cohorts (Turnbull, 2020). The underlying developmental mechanism is that students move from passive grade-recipients to proactive feedback-seekers, developing evaluative judgement and self-regulation as their relationship with assessment changes (Baartman & Quinlan, 2024). Students, like staff, absorb the initial costs of PLA in exchange for delayed benefits. They do so most successfully when the assessment architecture is stable enough for the transition to complete and that stability is what locally negotiated IPA implementations, dependent on individual goodwill and personal relationships, often cannot guarantee.

RQ 3. Management strategies and governance: mechanisms of sustained transition

3.1 Reframing the transition problem: from cultural to structural

The literature on PLA implementation acknowledges structural barriers like restrictive policies, governance frameworks and resource constraints, but tends to respond to them with cultural and psychological solutions. The main recommendations in the literature are professional development to improve assessment literacy (Charlton & Newsham-West, 2024b), incremental rollout to manage staff anxiety (Medina et al., 2013; Ryan & Judd, 2022) and leadership support to build buy-in (Jessop et al., 2012). All of these address the symptoms of the structural problem rather than its cause and as such the diagnosis and the prescription are mismatched. Even elaborate responses like institution-wide curriculum mapping systems and assessment cycle infrastructure developed in response to accreditation requirements, have been shown to have insufficient evidence of impact on student learning outcomes (Alyasin et al., 2023).

The more precise framing is structural, and it turns on what we term the cost-benefit asymmetry of PLA implementation. The Nottingham Civil Engineering redesign illustrates this directly: the costs of transition were front-loaded and immediate, curricula redesigned, assessment schedules rebuilt, marking uncertainty absorbed, while the benefits took years to materialise, with exam papers eventually reduced from nearly fifty to twenty-six and first-class outcomes rising from 28% to 39% (Turnbull, 2020). The people who bore those upfront costs were not the same people who received the downstream benefits.

This pattern is not unique to Nottingham: the benefits of PLA such as reduced marking loads (Schwegler et al., 2024), better student outcomes (Turnbull, 2020), improved graduate employability (Kotnour et al., 2022; Tree et al., 2025) are real and documented, but they arrive later, and often to different cohorts of staff and students than those who took the brunt of the initial disruption (Turnbull, 2020; Schwegler et al., 2024). This temporal and personal asymmetry is not a problem of attitude or communication; it is a structural feature of how implementation costs and returns are distributed across institutions, time, and people. The asymmetry was navigated rather than resolved at Nottingham, and the conditions that made navigation possible there are not universally available. Mechanisms that work durably are those that redistribute this asymmetry structurally, through protected time allocations (Turnbull, 2020; Tree et al., 2025), pass/fail grading that removes the marking overhead of numerical precision (Ryan & Judd, 2022), dedicated funding for curriculum redesign (Turnbull, 2020), and progression committees that distribute the burden of high-stakes decisions across a collective rather than placing it on individual staff (Baartman et al., 2022; van der Vleuten et al., 2012), rather than asking individuals to absorb it on goodwill and institutional faith and that is what the evidence from both model types, examined in the following subsections, consistently shows.

3.2 Programmatic assessment: structural redesign rather than change management

Mechanisms that sustain the transition to programmatic assessment point consistently to infrastructure rather than persuasion working by redistributing implementation costs structurally rather than asking individuals to absorb them. Incremental rollout is effective because it allows governance infrastructure like committees, portfolio systems, and the shared judgement frameworks that allow multiple assessors to make consistent progression decisions to develop progressively before the next phase is added (Medina et al., 2013; Ryan & Judd, 2022), not primarily because it manages anxiety. Each phase creates the structural preconditions for the next. The expert progression committee is the central mechanism: by distributing high-stakes decision-making across a collective governance structure rather than concentrating it on individual educators, it redistributes both the burden and the accountability of progression decisions institutionally (Baartman et al., 2022; van der Vleuten et al., 2012). Pass/fail grading works by the same logic applied to stakes rather than decisions, redistributing assessment risk across time rather than concentrating it in single high-stakes moments, with documented reductions in student anxiety and increases in peer collaboration as a result (Ryan & Judd, 2022).

Where professional accreditation frameworks exist, as in engineering and allied health, the same redistributive logic applies to legitimacy: external authority for redesign partially substitutes for institutional mandate, raising the standing of the change above what individual academic preference alone can claim (Turnbull, 2020). Digital monitoring infrastructure designed specifically to support outcome tracking has been shown to improve the efficiency of data collection and visualisation, but usability gains of this kind do not substitute for the governance conditions that make programme-level decision-making possible (Asmaranto et al., 2025).

3.3 IPA: relational mechanisms and their structural ceiling

The IPA evidence documents a different set of mechanisms from those operated in programmatic assessment: faculty learning communities, programme audit processes such as TESTA, incremental department-level rollout, that shift individual and team practice (Schwegler et al., 2024; Jessop, 2023; Lilly & Cooper, 2021). The conditions that make these work are well-documented: collegial cooperation, education-focused leadership, and sustained professional support (Charlton & Newsham-West, 2024b). The limit of these mechanisms, however, is equally consistent across the IPA evidence base: relational approaches reach a structural ceiling where sustained progress requires institutional mandate rather than individual initiative. Faculty learning communities can shift practice within a department but cannot restructure the validation processes, credit frameworks, and quality assurance systems governing assessment at institutional level. TESTA can reveal assessment patterns and stimulate programme-level conversations but cannot mandate that institutions act. Incremental rollout can build local momentum but cannot protect that momentum when key individuals move on or institutional priorities shift. The most developed available model of what that mandate looks like is the Nottingham Curriculum Transformation project, dedicated roles, financial incentives, a shared curriculum design framework, and formal institution-wide engagement (Turnbull, 2020). It has not solved the architecture problem, but it illustrates what moving beyond locally negotiated workarounds actually requires.

3.4 Redistributing the cost-benefit asymmetry

Across both programmatic assessment and IPA, we argue that the mechanisms which durably sustain transition share a common structural logic: they redistribute the costs of implementation away from individuals. This is our interpretation but is grounded in a pattern which emerges consistently from the literature, most directly in Charlton and Newsham-West's (2024b) finding that the forces that obstruct IPA implementation were systemic while the forces enabling it depended on personal relationships and goodwill

a distribution that describes precisely the asymmetry we are identifying. The principle becomes clearest when programmatic assessment, with its governance infrastructure, progression committees, and longitudinal portfolio systems, is placed alongside IPA, which has emerged more organically.

In programmatic assessment, progression committees redistribute the burden of high-stakes decision-making from individual educators to collective governance structures. Pass/fail grading at the data-point level redistributes assessment stakes across time rather than concentrating them in single high-stakes moments. Incremental rollout distributes the implementation burden across phases rather than front-loading it onto a single cohort of staff. These are structural redesigns of who bears what cost at what moment, and they are what makes programmatic assessment institutionally sustainable in the contexts where it has been sustained.

In IPA, the most effective mechanisms work by the same logic but with more limited institutional resources. Dedicated time allocations redistribute the cost of curriculum redesign from individuals' discretionary effort to institutional investment. Financial incentives, as at Nottingham, redistribute the economic burden of participation from departments to central institutional funding. Professional accreditation requirements, as at Nottingham Civil Engineering, redistribute the legitimacy of the redesign from individual academic preference to external professional authority.

The practical implication for institutional leaders suggests that IPA cannot be sustained at scale through goodwill and relational persuasion alone: not because academics are unwilling, but because the cost-benefit asymmetry of implementation is structurally unreasonable to absorb on an individual basis. Sustainable transition requires institutions to identify, explicitly, which mechanisms will redistribute implementation costs and to invest in facilitating those mechanisms rather than simply encouraging individuals to manage through the transition period. The question is not 'how do we persuade staff to change?' but 'what structural changes would alter the cost-benefit calculation for the people being asked to change?'

Conclusion

Three Conclusions

The central finding of this work is that PLA implementation fails primarily because of institutional architecture rather than cultural resistance. The governance systems, quality assurance frameworks, and credit structures through which universities regulate assessment are designed to reproduce modular logic, and staff resistance, though real, is downstream of this more fundamental problem. Addressing the symptoms through cultural change management alone, while leaving the architecture intact, is insufficient.

The two principal models of PLA, programmatic assessment and IPA, are analytically distinct rather than interchangeable approaches to a shared problem. Both face the same institutional architecture problem, but with structurally different resources: programmatic assessment has been able to build parallel governance infrastructure in contexts where medical schools possessed the institutional standing to do so, while IPA in broader higher education relies on locally negotiated workarounds that are pragmatic but structurally fragile.

The mechanisms that have been found to sustain transition in both models work by redistributing the cost-benefit asymmetry of implementation rather than asking individuals to absorb it. The upfront costs of PLA fall on different people at different times than the downstream benefits, and this structural feature, more than any cultural disposition, explains why goodwill and incremental persuasion alone cannot sustain adoption at scale.

Implications for practice and policy

The practical implication is direct: institutions need to stop asking how to persuade staff to change and start asking what structural changes would alter the cost-benefit calculation for the people being asked to change. That means governance restructuring must precede or accompany cultural change programmes, not follow them. Validation documents, credit frameworks, and assessment policies need to be redesigned to recognise and require programme-level approaches before professional development and change management can be expected to take hold. The Nottingham Curriculum Transformation project represents the most developed available model of what this looks like in practice for IPA: dedicated roles, financial incentives, a shared curriculum design framework, and formal institution-wide engagement (Tree et al., 2025). Institutions considering IPA implementation should regard this not as an aspirational advanced case but as a minimum viable model for sustained adoption.

For those who fund, regulate, and quality-assure higher education, the implication is equally direct. The evidence reviewed here suggests that the disciplinary reach of current IPA advocacy considerably exceeds the disciplinary scope of the evidence base. The field needs systematic research in humanities, social sciences, and arts disciplines before confident claims about broad applicability can be made. Funding bodies and quality agencies that mandate or incentivise programme-level assessment without simultaneously restructuring the governance frameworks that make it impossible are asking institutions to do something the system is not yet built to support.

References

- Alyasin, A., Nasser, R., El Hajj, M., & Harb, H. (2023). Assessing learning outcomes in higher education: From practice to systematization. *TEM Journal*, 12(3), 159–1604. <https://doi.org/10.18421/TEM123-41>
- Asmaranto, N. A., Kusumasari, T. F., & Alam, E. N. (2025). A dashboard system for enhancing learning outcome monitoring and curriculum management under the outcome-based education framework. *Proceedings – International Seminar on Intelligent Technology and its Applications, ISITIA*. 661-666 <https://www.doi.org/10.1109/ISITIA66279.2025.11137419>
- Baartman, L., van Schilt-Mol, T., & van der Vleuten, C. (2022). Programmatic assessment design choices in nine programs in higher education. *Frontiers in Education*, 7. <https://doi.org/10.3389/feduc.2022.931980>
- Baartman, L. K. J., & Quinlan, K. M. (2024). Assessment and feedback in higher education reimaged: using programmatic assessment to transform higher education. *Perspectives: Policy and Practice in Higher Education*, 28(2), 57–67. <https://doi.org/10.1080/13603108.2023.2283118>
- Barker, T. H., Habibi, N., Aromataris, E., Stone, J. C., Leonardi-Bee, J., Sears, K., Hasanoff, S., Klugar, M., Tufanaru, C., Moola, S., & Munn, Z. (2024). The revised JBI critical appraisal tool for the assessment of risk of bias for quasi-experimental studies. *JBI Evidence Synthesis*, 22(3), 378–388. <https://doi.org/10.11124/jbies-23-00268>
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32(3), 347–364. <https://doi.org/10.1007/BF00138871>
- Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, 22(2), 151–167. <https://doi.org/10.1080/713695728>
- Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: the challenge of design. *Assessment & Evaluation in Higher Education*, 38(6), 698–712. <https://doi.org/10.1080/02602938.2012.691462>
- Boud, D., & Soler, R. (2016). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, 41(3), 400–413. <https://doi.org/10.1080/02602938.2015.1018133>
- Bowe, C. M., & Armstrong, E. (2017). Assessment for systems learning: A holistic assessment framework to support decision making across the medical education continuum. *Academic Medicine*, 92(5), 585–592. <https://doi.org/10.1097/ACM.0000000000001321>

Brown, G. A., Bull, J., & Pendlebury, M. (2013). Assessing student learning in higher education. Taylor & Francis.

Charlton, N., & Newsham-West, R. (2024a). A conceptual model for program-level assessment. *Higher Education Research & Development*, 43(8), 1721–1736.
<https://doi.org/10.1080/07294360.2024.2364094>

Charlton, N., & Newsham-West, R. (2024b). Enablers and barriers to program-level assessment planning. *Higher Education Research & Development*, 43(5), 1074–1088.
<https://doi.org/10.1080/07294360.2024.2307933>

Charlton, N., Weir, K., & Newsham-West, R. (2022). Assessment planning at the program-level: a higher education policy review in Australia. *Assessment & Evaluation in Higher Education*, 47(8), 1475–1488.
<https://doi.org/10.1080/02602938.2022.2061911>

Clarke, J. L., & Boud, D. (2018). Refocusing portfolio assessment: Curating for feedback and portrayal. *Innovations in Education and Teaching International*, 55(4), 479–486.
<https://doi.org/10.1080/14703297.2016.1250664>

Crockett, J. L., Fu, A. S., Greenwood, J. L., & John, M. A. (2018). Integrated sociology program assessment: Inclusion of a senior portfolio graduation requirement. *Teaching Sociology*, 46(1), 34–43. <https://doi.org/10.1177/0092055X17726833>

Dukhi, N., Southwood, S., & Srinivas, S. C. (2014). Evaluating students' experience of an integrated assessment: A case study in health promotion. *Indian Journal of Pharmaceutical Education and Research*, 48(3), 1–5.
<https://doi.org/10.5530/ijper.48.3.1>

Eva, K. W., Bordage, G., Campbell, C., Galbraith, R., Ginsburg, S., Holmboe, E., & Regehr, G. (2016). Towards a program of assessment for health professionals: from training into practice. *Advances in Health Sciences Education*, 21(4), 897–913.
<https://doi.org/10.1007/s10459-015-9653-6>

Fielding, D. W., & Regehr, G. (2017). A call for an integrated program of assessment. *American Journal of Pharmaceutical Education*, 81(4), Article 77.
<https://doi.org/10.5688/ajpe81477>

Gibbs, G., & Dunbar-Goddet, H. (2009). Characterising programme-level assessment environments that support learning. *Assessment & Evaluation in Higher Education*, 34(4), 481–489. <https://doi.org/10.1080/02602930802071114>

Gorra, A., Lazarevski, S., & Campbell, J. (2007). An account of the use of integrated assessment for students in the area of databases at level 2. [Conference proceedings].

Harland, T., McLean, A., Wass, R., Miller, E., & Sim, K. N. (2015). An assessment arms race and its fallout: high-stakes grading and the case for slow scholarship. *Assessment & Evaluation in Higher Education*, 40(4), 528–541.

<https://doi.org/10.1080/02602938.2014.931927>

Hartley, P., & Whitfield, R. (2012). Programme Assessment Strategies (PASS) final report. Higher Education Academy.

https://www.brad.ac.uk/pass/about/PASS_final_report.pdf

Heeneman, S., de Jong, L. H., Dawson, L. J., Wilkinson, T. J., Ryan, A., Tait, G. R., Rice, N., Torre, D., Freeman, A., & van der Vleuten, C. P. (2021). Ottawa 2020 consensus statement for programmatic assessment 2: Implementation and practice. *Medical Teacher*, 43(10), 1139–1148. <https://doi.org/10.1080/0142159X.2021.1957088>

Hughes, G., Smith, H., & Creese, B. (2015). Not seeing the wood for the trees: developing a feedback analysis tool to explore feed forward in modularised programmes. *Assessment & Evaluation in Higher Education*, 40(8), 1079–1094.

<https://doi.org/10.1080/02602938.2014.969193>

Jessop, T. (2023). Student agency and engagement: Transforming assessment and feedback in higher education. Routledge. <https://doi.org/10.4324/9780429347962>

Jessop, T., McNab, N., & Gubby, L. (2012). Mind the gap: An analysis of how quality assurance processes influence programme assessment patterns. *Active Learning in Higher Education*, 13(2), 143–154. <https://doi.org/10.1177/1469787412441285>

Jessop, T., & Tomas, C. (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>

Jimenez, Y. A., Punch, A., Lewis, S. J., & Reed, W. (2022). Teaching evidence-based practice: Case study of an integrated assessment task for diagnostic radiography students. *Journal of Medical Imaging and Radiation Sciences*, 53(3), 341–346.

<https://doi.org/10.1016/j.jmir.2022.05.005>

Kotnour, T. G., Farr, J. V., & Vergopia, C. (2022). An integrated approach for assessing graduate engineering management programs. *International Journal of Engineering Education*, 38(4), 1092–1105.

Kovaichelvan, V., & King, C. S. (2020). Program assessment through product based learning in undergraduate engineering programmes in India. ASEE Annual Conference and Exposition, Conference Proceedings. <https://doi.org/10.18260/1-2--35093>

Lilly, T., & Cooper, R. (2021). Getting to eportfolios through assessment: A process model for integrating eportfolio pedagogical concepts through training for program assessment. *Journal of Higher Education Theory and Practice*, 21(7), 44–55.

<https://doi.org/10.33423/JHETP.V21I7.4485>

Lockwood, C., Munn, Z., & Porritt, K. (2015). Qualitative research synthesis: methodological guidance for systematic reviewers utilizing meta-aggregation. *International Journal of Evidence Based Healthcare*, 13(3), 179–187.

<https://doi.org/10.1097/xeb.0000000000000062>

Medina, M. S., Britton, M. L., Letassy, N. A., Dennis, V., & Draugalis, J. R. (2013). Incremental development of an integrated assessment method for the professional curriculum. *American Journal of Pharmaceutical Education*, 77(6), Article 122.

<https://doi.org/10.5688/ajpe776122>

Naeem, M., Ozuem, W., Howell, K., & Ranfagni, S. (2023). A Step-by-Step Process of Thematic Analysis to Develop a Conceptual Model in Qualitative Research. *International Journal of Qualitative Methods*, 22.

<https://doi.org/10.1177/16094069231205789>

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>

Pye, R. E., Gnagi, L. K., Chuah, A. B. Q., & Chung, K. L. (2025). High stakes in the assessment arms race: A mixed-methods analysis of reducing assessment in an undergraduate psychology programme. *Psychology Learning & Teaching*, 24(2), 195–213. <https://doi.org/10.1177/14757257251336295>

Rust, C. (2002). The impact of assessment on student learning: How can the research literature practically help to inform the development of departmental assessment strategies and learner-centred assessment practices? *Active Learning in Higher Education*, 3(2), 145–158. <https://doi.org/10.1177/1469787402003002004>

Rutherford, S. M. (2019). 'Flying the nest': An analysis of the development of self-regulated learning during the transition to higher education [Doctoral dissertation, University of Reading].

Ryan, A., & Judd, T. (2022). From traditional to programmatic assessment in three (not so) easy steps. *Education Sciences*, 12(7), 487.

<https://doi.org/10.3390/educsci12070487>

Schuwirth, L. W. T., & Van der Vleuten, C. P. M. (2011). Programmatic assessment: From assessment of learning to assessment for learning. *Medical Teacher*, 33(6), 478–485. <https://doi.org/10.3109/0142159X.2011.565828>

Schwegler, A. F., Shell, M. D., & Clark, D. A. (2024). Using Guidelines 3.0 to support program and course learning outcome alignment. *Scholarship of Teaching and Learning in Psychology*, 11(1), 65–80. <https://doi.org/10.1037/stl0000414>

Tree, D., Harvey, A., & Rand-Weaver, M. (2025). Integrated Programme Assessment: 10 years on. *Open Scholarship of Teaching and Learning*, 4(1), 166–180. <https://doi.org/10.56230/osotl.128>

Turnbull, B. (2020). An industry-inspired civil engineering curriculum. *Proceedings of the Institution of Civil Engineers – Civil Engineering*, 173(2), 91–95. <https://doi.org/10.1680/jcien.19.00012>

van der Vleuten, C. P. M., Schuwirth, L. W. T., Driessen, E. W., Dijkstra, J., Tigelaar, D., Baartman, L. K. J., & van Tartwijk, J. (2012). A model for programmatic assessment fit for purpose. *Medical Teacher*, 34(3), 205–214. <https://doi.org/10.3109/0142159X.2012.652239>

Vanderheide, R. A., & Walkington, J. (2009). Maximised learning through integrated assessment: Evidenced through nursing practice. *Nurse Education in Practice*, 9(6), 351–355. <https://doi.org/10.1016/j.nepr.2008.10.010>

Walker, S. (2025). *Trends in assessment in higher education: considerations for policy and practice*. Jisc. Retrieved from <https://repository.jisc.ac.uk>

Figure 1: Contrasting assessment architectures in higher education.

(A) Modular assessment, in which teaching subjects are assessed through separate coursework and examinations.

(B) A hybrid model combining module-based and programme-level assessment, with some assessments remaining module-specific and others designed to integrate learning across modules.

(C) Integrated programme assessment (IPA), in which assessment is organised entirely at programme level through integrative tasks that draw on learning across modules.

(D) Programmatic assessment, developed primarily in medical education, which conceptualises assessment as a longitudinal system of connected low-stakes assessment points feeding into high-stakes progression decisions.

Figure 2: PRISMA flowchart detailing the study search and selection process applied during the systematic review. The flowchart depicts the process of identification, screening, eligibility assessment, and inclusion of studies. The initial search across three databases (Scopus, Web of Science and ProQuest) resulted in 632 records, with one additional article identified through other sources. After the removal of duplicates ($n = 95$) and using filters on the databases themselves to exclude articles based on the predefined criteria (shown in bold in Table 2; $n = 306$). A total of 232 records were screened by the review team evaluating the title and abstract, again using the exclusion criteria listed in Table 2. Full text reports were sought for 63 articles, with 58 accessible. These were assessed against the same exclusion criteria through a detailed reading of the article. Of these, 33 were excluded, resulting in 25 studies that met the inclusion criteria and were incorporated into the final analysis.

Table 1. Scopus, and ProQuest aligned search strings, with any limits noted.

Table 2. Exclusion criteria, in order of application, used to screen initial literature search results. Criteria in bold were applied during primary screening using database filters. All of the criteria were applied in secondary and tertiary screening rounds by the review team.