

> Next Generation Assessment



Online
Education
Services

Reimagining online learning

Key Messages

- > Student assessment is a multibillion dollar pillar of higher education
- > Innovation is needed and feasible to provoke step-change reform
- > Next-generation assessment will create substantial fresh value for all

> Contents

The Shape of Things to Come	3
Stand Back Look Ahead	5
Creating Perspectives	7
Almost Anything Goes	10
Assessment Architecture	12
Steps Ahead	16

> Acknowledgements

This research briefing was prepared by Professor Hamish Coates (Tsinghua University) in collaboration with experts at Online Education Services (OES). Thank you to Kenneth Moore (University of Melbourne) for earlier research assistance. With special thanks to many experts and stakeholders who have contributed to this briefing.

> Publishing/copyright

First published 2018 by Online Education Services Pty Ltd ("OES")
Copyright © OES 2018. All rights reserved.

Without limiting the rights under copyright reserved above, no part of this publication may be reproduced, stored in or introduced into a database and retrieval system or transmitted in any form or any means (electronic, mechanical, photocopying, recording or otherwise) without the prior written permission of OES.

Contributor credits: OES acknowledges the contribution of Professor Hamish Coates (Tsinghua University), and Kenneth Moore (University of Melbourne) for earlier research assistance. Special thanks to the many OES staff members who also contributed to this paper.

> The Shape of Things to Come

When people born this year enter university they will experience very different assessment practices to those in widespread use today. They will experience 'next generation assessment'. Assessment will be woven invisibly into education experiences, almost indistinguishable from 'curriculum' and 'teaching' and 'support'. It will sculpt each learning moment with challenges and guides that help each person reach their potential. Resources will be produced by education engineers who share a scientific commitment to their profession. Assessment will yield tailored insights for a range of audiences that help people move seamlessly through study and work.

This briefing presents outcomes from a project designing an architecture for next generation assessment. It invites people with vested interests in higher education to recognise current problems and imagine the shape of things to come (**Figure 1**). In this briefing 'assessment' is interpreted broadly as involving the measurement, reporting and interpretation of student learning and development. This includes what are sometimes called 'formative assessment' and 'summative assessment' – a distinction very likely to fade in years to come as focus shifts to people and their learning rather than the locus and rationale of supply.

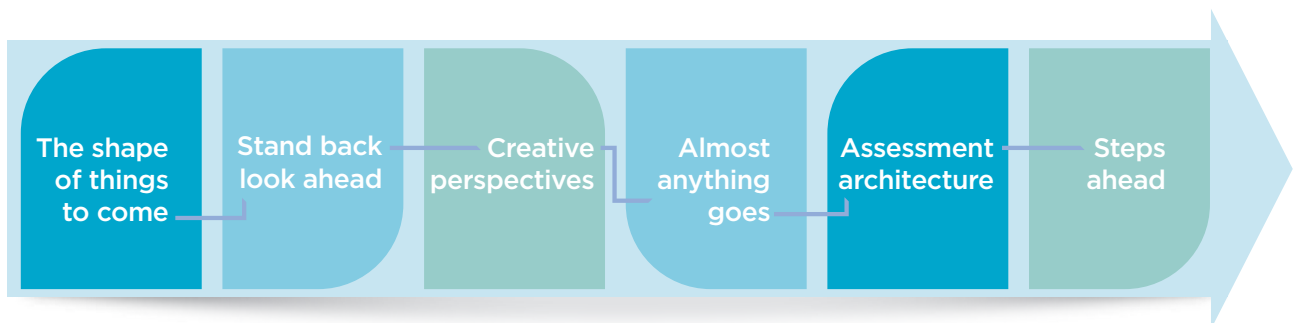


Figure 1: Flow of this briefing

Assessment matters enormously to higher education. It bookends who 'gets in' and who 'gets out'. It is the distilled essence of what is taught. It drives learners' engagement and signposts their achievements. It gatekeeps education standards. It distinguishes different programs and institutions.

Assessment has yet to have its transformational moment despite other major disruptions to higher education. Hypertravel has augmented the swirl of students, academics and universities around the world. Online learning has transformed curriculum and teaching in ways that could be mild compared with the artificial transmutations to come. Student numbers have grown and diversified. New fields have sprouted and matured. Yet much assessment is being done today much as it was a century ago.

But as with any important thing there are cogent reasons for improvement. There is value in advancing assessment in the spirit of continuous improvement. There are strategic institutional rationales for finding innovative ways to assess student learning. Graduate employer and business concerns about education standards ultimately fall back to concerns about assessment. Assessment is a fulcrum for enhancing student engagement and retention. At the same time, by doing assessment better and cheaper there is enormous educational and financial value to be found for institutions, faculty, students and governments. Producing more cogent data on outcomes would yield broader dividends by proving economic and social returns from education—important additions to other information about quality.

“Business as usual’ is unlikely to prevail ... there is a need for smart workarounds”

There have been many attempts to transform assessment. It is important to recognise the innovative but local efforts of millions of academics. Larger signature initiatives have involved imposing qualification frameworks, cross-national benchmarking projects, discipline-specific reforms, the tuning of curriculum structures, training scorers, and implementing standardised assessments. Such initiatives have registered various advances, yet none have tipped the scales and spurred pervasive new norms or practices.

Deep analysis of higher education exposes interlocking forces hindering change. Though it is difficult to be conclusive it appears several key forces are at play (Figure 2). Many of these hindering forces seem exogenous to assessment itself. Such complexity highlights the significance of assessment. It also signals that ‘business as usual’ is unlikely to prevail. There is a need for smart workarounds. Change is likely to be provoked by capabilities established through new partnerships and infrastructure.

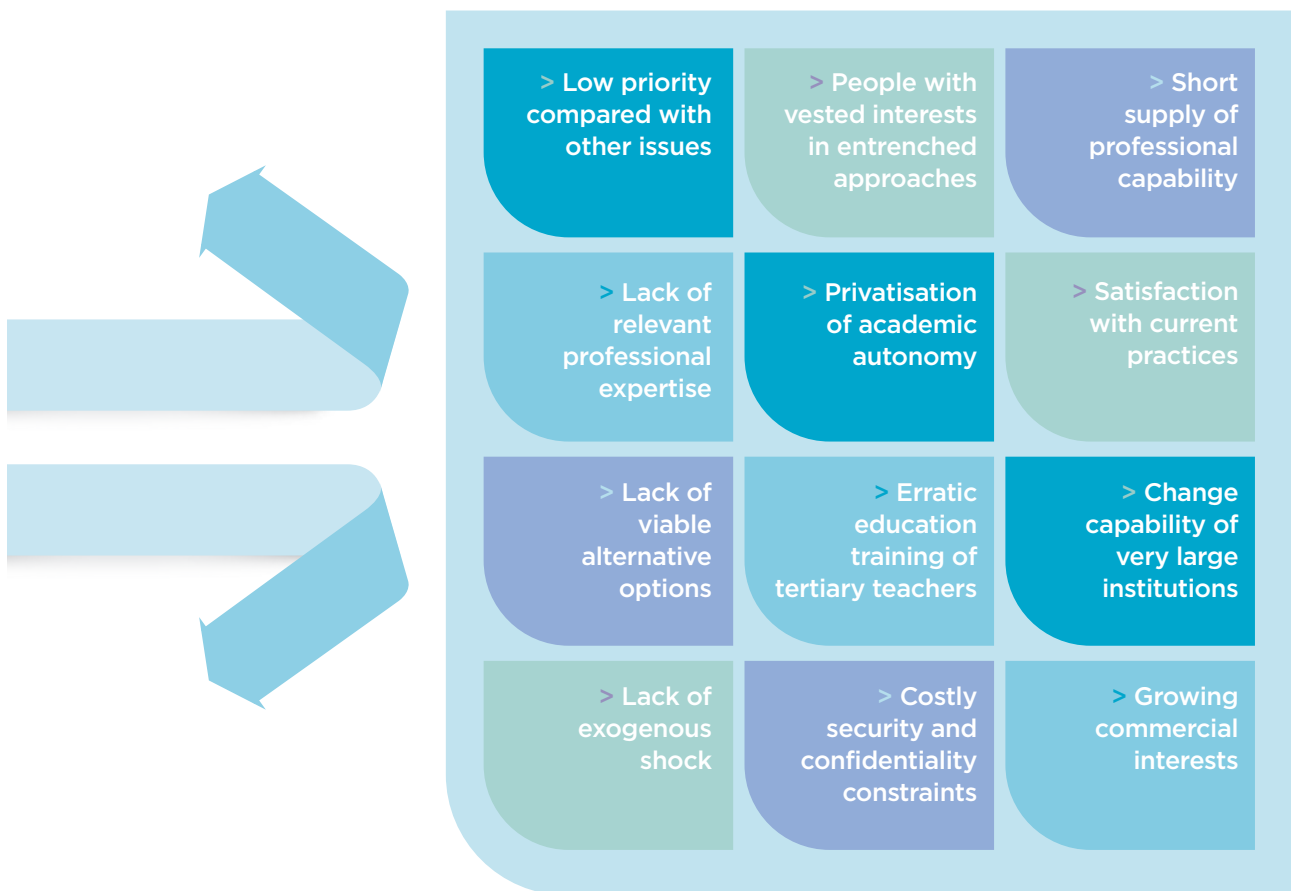


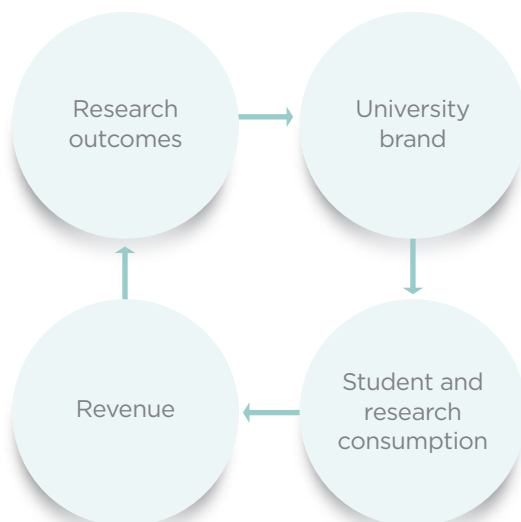
Figure 2: Change blockers

> Stand Back Look Ahead

Almost anyone who works in higher education knows that now is not business as usual.¹ Demand keeps heating. Global student numbers could reach 265 million by 2025, up from 100 million in 2009. The number of people participating in Australian higher education has risen by 40 per cent over the last decade. Economies are booming and maturing, spurring economic thirst for more educated workers. The market for credentials accelerates with the global brain hunt, both because of and in spite of AI/AR rhetoric. Longer lives means longer careers and more retraining. Yet supply is fraught. Traditional education models have been stretched to bursting. Legacy workforce arrangements enjoyed by privileged elites overshadow the realities of contemporary academic roles and structures. Directly and through government, the public is asserting new accountabilities and financial pressures. Even at the most personal and local levels, higher education keeps changing in complex synchronisation with non-ignorable global developments.

Now is a great time to invent future forms of higher education. So, what helpful thinking can create fresh value for the core education services that drive most of the rest of the world's higher education institutions? One temptation, particularly in any of the world's more entrepreneurial contexts, is to crave financial surplus from consuming students. Another option, healthier and even more prudent, is to work with students to co-create alpha education returns. Great value is to be found by those education innovators who create new education highs rather than race to any financial bottoms. Near-sighted financial trading shapes very few parts of education, which better resembles an ultimate game of value investment. Rather than parlay research trophies into education profits, and vice versa, a more coherent logic would see education success as creating education value, which in turn spurs education investment, which then leads to expanded contributions, which inspires further successes (Figure 3). This new kind of thinking carries enormous potential to shine light which strengthens people and entire education systems.

> Prevailing logic



> New logic

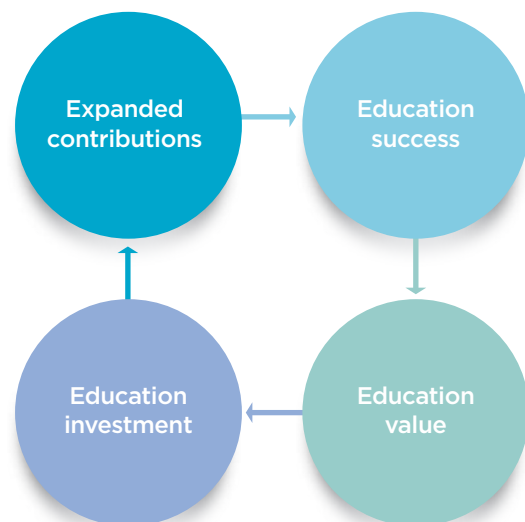


Figure 3: New logic for next-generation higher education

1 | Coates, H. (2017). *The Market for Learning: Leading transparent higher education*. Dordrecht: Springer.

Reforming the assessment of student learning is surely a plinth in future growth. Prognosing the future of assessment is fascinating but carries scary implications if not led well. Changing products and services means reconfiguring legacy structures, roles, materials, interactions and even ambitions. In existing contexts assessment is woven into close collaboration between academics and students. In future settings assessment dissolves into the technologically mediated interplay between learners and much broader learning systems. This has direct implications for people, infrastructure, and institutions. But the implications are not immediate, despite technozealot rhetoric. Change is paced not just by software advance but also and perhaps more powerfully by broader transformations of anxious people, complex institutions, and massive systems. The world is littered by IT solutions broken from neglect of people.

To position progress, the evolution of assessment practice can be framed in three eras. To exemplify these **Figure 4** highlights a few shifting practices in the transition from traditional through stretched to next generation practices. The last few decades have been expansionary though not transformative in nature. Conservative reckoning implies that the number of pieces of assessment each year in Australian higher education has risen from around seven

to nearly 30 million over the last twenty years. With no change in production function this implies a quadrupling in recurrent annual costs to more than \$400,000,000 for marking alone, leaving aside all of the other large costs for assessment development, administrative and support staff, capital and intermediaries (see **Figure 6**). All up, including indirect costs, it is easy to see how largely unreformed assessment practices might be costing close to a billion dollars annually. Education practices common in traditional settings may be wonderful but hard to scale. These legacy practices have been stretched taut and patched for bigger delivery as higher education has expanded. The architecture below offers a gangplank for porting to truly enriched next generation practice which has the properties pitched at the outset of this briefing.

After decades if not centuries of application, aspects of assessment still in widespread use today are almost surely running out of puff. Certain zombie practices have been propped up well beyond maturity into various kinds of sustained obsolescence. The S-curve has been stagnant for some time. Given this, and the delivery of new technologies now is ripe to ferment genuine improvements in productivity. Any open-minded glance at assessment reveals rocks under which there are likely to be step-change solutions.

	Traditional > 1990s and before	Stretched > 1990s to 2020	Next generation > 2020s and after
Authority	University	University or regulator	Shared
Production	Solo academics	Academic teams	Co-creation
Format	Paper	Paper and online	Online
Location	Campus	Campus and online	Online
Implementation	Universities	Universities	Specialists
Scoring	Solo academics	Moderated practice	Automated
Reporting	Generic	Contextualised	Customised

Figure 4: Three eras of assessment

> Creating Perspectives

Design thinking presents a useful epistemology and methodology for advancing assessment. Design thinking is an interactive five-step process.¹ The first stage ‘empathising’ means defining the problem and exploring relevant human contexts—what is the problem, and who cares? ‘Definition’ then involves determining and defining the things that are important—creating a point of view based on needs and insights. ‘Ideation’ involves brainstorming possible futures—generating as many wild ideas as possible. ‘Prototyping’ is about building the new things and approaches enough to convey nature and value. And ‘testing’ means looking at whether the things and approaches work, typically with reference to initial audience and needs. These steps might be bundled into three broader phases such as inspiration (understanding and observing), ideation (thinking, prototyping and testing) and implementation (storytelling, piloting and scaling).

Management research reveals that design thinking helps organisations and professionals out-perform.² It steps around, above and beyond traditional scientific and educational thinking, encouraging exploration and experimentation rather than conservative incrementation.

It helps deliver a product rather than just a set of thoughts. It charts course towards novel rather than nudging existing thinking. It encourages creative controversy rather than incumbent functionality

Figure 5 presents the overall redesign logic deployed in this innovation. With its tentacles tucked deep around the roots of education it is no small question to ask where assessment reform energy should best be funnelled. What investments will provoke the most change? Guidance can be found by framing then deconstructing assessment, diagnosing problems, identifying improvements, and advancing an engineered solution.

Design thinking helps harness eclectic strands of thought and channel these into assessment inventions. Doing different things needs new educational thinking, new corporate structures, new capability, and new partnerships. Weaving together several lines of education and business thinking contribute very helpful perspectives. These include Academic Course Redesign, Evidence Centred Design, Third Generation Assessment, Business Process Re-engineering, and Value Chain Analysis. As design thinking conveys, sensitive analysis of key contexts is also required.

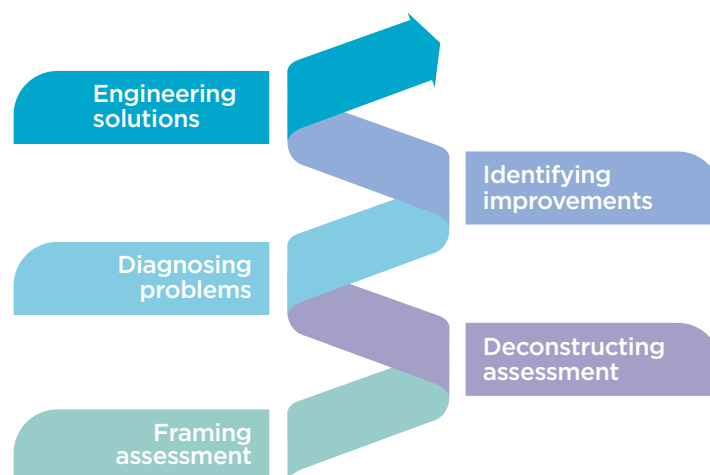


Figure 5: Assessment redesign logic

1 | Design Management Institute (www.dmi.org)

2 | Edson, J., Kouyoumjian, G. & Sheppard, B. (2017). *More than a Feeling: Ten design practices to deliver business value*. New York: McKinsey & Company.

“Evidence Centred Design (ECD) provides a holistic and arguably the best platform for assuring the development, review and improvement of assessment. ECD ideas, resources and techniques service foundations for next generation assessment”

Launched by Carol Twigg at the United States National Center for Academic Transformation (NCAT), Academic Course Redesign refers to the process of analysing and improving education outcomes while reducing cost, making as much use of IT as possible.¹ Course Redesign was developed to assist academic coordinators at traditional universities to reduce course-level costs and improve outcomes, however the ideas are readily applicable to the subject/unit level, and to the assessment function. In essence, it involves selecting a redesign model, defining student outcomes and appropriate assessment methods, and specifying detailed cost reduction strategies. Over several years NCAT has developed very specific resources and approaches for the application of these approaches. Of most use to the current project is the idea that commercial logic can be applied to existing/legacy education practices to improve productivity.

Advanced by Robert Mislevy at the United States Educational Testing Service (ETS), Evidence Centered Design (ECD) articulates a scientific framework for designing, producing and delivering educational assessments with respect to evidence requirements.² Drawing from disciplines like psychology, IT and policy studies, the framework includes models and processes that ensure assessment is sound and effective. ECD includes modelling of the domain/topic being assessed. It includes mapping important facets of assessment such as students, the evidence required, optimal task specifications, implementation requirements, and delivery specifications. ECD provides a holistic and arguably the best platform for assuring the development, review and improvement of assessment. ECD ideas, resources and techniques service foundations for next generation assessment.

The term ‘next generation assessment’ flows from foresight research by Randy Bennett, also at ETS.³ While ECD establishes a framework for structuring work on assessment, Bennett’s work contributes an historical lens. The evolutionary story is about how IT can, though often may not, make assessment better. In a nutshell, ‘first-generation assessment’ differs only marginally from paper-based assessment. This type of practice simply involves loading paper-derived materials onto online systems, and deploying them in relatively standard ways. ‘Second-generation assessment’ invokes the use of IT-derived innovations in educational practice, though often without any more advanced educational underpinnings, or without a fulsome marriage between educational and technological considerations. ‘Third-generation assessment’ involves the formation and promulgation of sophisticated educational and technological reasoning. IT is designed around new or rejuvenated educational ideas/practices, and educational practices are transformed around contemporary platforms. These perspectives play a very important motivating and shaping role in the current design. They help expose first- or second-generation practices, and set principles and rationales for launching the next generation.

Education-flavoured thinking is not itself sufficient to launch next generation assessment. Other organisational settings, can be the inspiration for alternative assessment approaches and should be taken into account. Business Process Re-engineering helps analyse and improve organisational workflows and processes.⁴ BPR is a management strategy that involves analysing an organisation’s mission, processes, decisions, information and technology. From this, organisations or their parts can then be redesigned to improve processes and ensure mission alignment. BPR can provoke big restructures. In the current design work BPR is focused on assessment sub-processes.

1 | National Center for Academic Transformation (www.thencat.org)

2 | Mislevy, R.J., Almond, R.G. & Lukas, J.F. (2003). *A Brief Introduction to Evidence-centered Design*. Princeton: Educational Testing Service.

3 | Bennett, R. 2015. The Changing Nature of Educational Assessment. *Review of Research in Education*, 39(1), 370-407.

4 | Davenport, T. (1993). *Process Innovation: Reengineering work through information technology*. Boston: Harvard Business School Press.

Planning	Developing	Implementing	Analysing	Reporting
> Governance > Leadership > Management	> Mapping resources > Specifying outcomes > Selecting formats > Drafting materials > Qualitative review > Quantitative review > Material production	> Designing administration > Organising facilities > Managing students > Administering assessments > Resolving problems	> Collating results > Marking > Producing data > Cross-validating results	> Producing grades > Analysing and commenting > Reporting > Reviewing and improving

Figure 6: Assessment phases and activities

Value Chain Analysis, distilled originally by Michael Porter, is a helpful management technique.¹ The 'value chain' is comprised of activities performed to convert 'inputs' into 'outputs'. VCA involves studying these activities and identifying how to reduce costs and improve quality—that is, how to improve productivity. This management tool is useful for mapping assessment processes, identifying competitive advantages (through better/differentiated quality) and disadvantages (bottlenecks, inefficiencies, etc.), and identifying productivity improvement opportunities. VCA can be conducted in different ways to suit different needs. Typically, it involves reviewing primary activities (e.g. inbound logistics, operations, outbound logistics, marketing, service) and enablers (e.g. infrastructure, human resources, technology and procurement). These primary activities and enablers can all be phrased in assessment/educational rather than commercial terms. **Figure 6** models this in terms of key assessment phases and activities. This model draws from professional measurement science research which has produced generalisable insights into assessment. Borrowing these mechanisms avoids myopia by situating higher education in the broader world, and helps garner the most scientifically relevant insights. The model provides a compelling lens for reviewing how assessment creates value for people.

Assessment involves real people and places, not just educational and organisational thinking. To gain traction any design of future practice must have legs as well as brain, and also heart. It must take account of people's interests and skills, organisational contexts, required change capabilities, regulatory settings, professional configurations, and more tacit cultures. There is no magic or secret to reading, manoeuvring within or rejuvenating communities or situations. Too little regard or too much deference spells stagnation which yields failure. Finding the sweet spot is difficult but critical for success.

These theories and perspectives are not adopted and applied in simple or slavish ways, but rather are drawn in various ways into the design. Such integration is reasonably distinctive. Assessment research and consultation that is international in scope has highlighted the need for and value of these various developments. Despite marked educational and commercial advantages there has been little attempt to date to weave these together to form new ways for doing business.

1 | Porter, M.E. (1985). *The Competitive Advantage: Creating and sustaining superior performance*. New York: Free Press.

> Almost Anything Goes

Birds assemble foraged materials into crafty infrastructure to create new life. Rummaging around current assessment practice reveals much material for future innovation. And there is much to be observed. Deep-diving into everyday practice in even very advanced and accredited contemporary contexts conveys that anything goes, almost. A long list of insights was produced by three months' observation of existing education practice across three diverse discipline areas. This observation involved meetings, conversations, reviews, and prototyping. It pinpointed opportunities for improvement and innovation, and the opportunity to trial prototypes of next generation assessment architecture.

"As much assessment evolves from existing materials there can be large and undocumented task variants in use, a 'flowering of practice' that creates inefficiencies and quality uncertainties."

At the outset, it is helpful to characterise features of traditional assessment practice. This includes higher education teachers working alone inside single institutions using technically non-validated materials. It involves assessment with an arbitrary mapping to curriculum, which are focused on single subjects, delivered using dated methods, scored normatively by uncalibrated markers, adjusted to fit percentile distributions, and reported using grades and thin feedback. There are often evident though ignored conflicts of interest, whereby the same person has full charge over the teaching /curriculum/assessment triad, and over all facets of assessment such as design, delivery, analysis, and reporting. Not everything is bad in such practice, and not all practices are used all at once or all of the time. Much more informality is feasible in elite settings. But each of these points signposts practices likely to fade in coming years.

The alignment between learning outcomes and assessment tasks can be challenging. Learning outcomes can be defined in very broad ways, partly to cover over invariably but not necessarily indeterminate links with lectures, tutorials and assessment tasks. Tasks for units can be designed and developed in holistic or impressionistic rather than scientific ways, leading to unclear or suboptimal configurations, creating opportunity to better integrate and align assessment with the other facets of education. Next generation assessment should have tasks which are aligned with articulated learning outcomes.

Assessment tasks may not have been defined and parameterised and developed in clear and coherent ways. Even common tasks and practices like groupwork, exams and essays may not be defined in any technical sense. The weightings given to pieces of assessment within a unit can be arbitrary and not be fully explained by factors such as sequencing, length or task rationale. The length of tasks, in terms of word lengths and presentation timings, often appear to flow from prior practices as much as considered educational or commercial perspectives. Instructions can vary across tasks as they are produced by different teaching/design personnel, creating inefficiencies and potential confusion and unfairness for students. As much assessment evolves from existing materials there can be large and undocumented task variants in use, a 'flowering of practice' that creates inefficiencies and quality uncertainties. There may even be variations in the nature and quality of tasks even of a similar type. Given the lack of broader specification, basic technological infrastructure can frame practice. Inadvertently, this also means that tasks may not make full use of online options/functions. Next generation assessment should involve tasks which are fully defined and specified.

“While it is core to education, people are not often employed solely as assessment specialists who take overall expert responsibility for the quality and productivity of assessment.”

Assessment may not be delivered in ways that enrich each student's experience. Some assessment may be only accidentally related to the education experience, which partly stems from lack of alignment with other facts of the unit, but mostly from an instructor- rather than student-oriented view of education. There can be batch processes in play, signalling opportunities to individualise and to make more use of real case studies, simulation experiences, or make feedback more aligned with people's interests and aspirations. There can be known 'hassles' arising from the administration and experience of group work and presentations which may not have been addressed due to concerns about fidelity with past practices. Collaborative work might be advanced in more effective and enjoyable ways by minimising administrative constraints around joint synchronous work, and instead deferring to more sophisticated asynchronous interactions which make use of collaborative features like meta-data and student background information. Getting the pitch of assessment might be difficult with diversifying student cohorts. Certain assessments may make assumptions about student aptitude and preparedness that are no longer feasible given larger and more diverse student bodies. Next generation assessment should be driven by supporting, challenging and enriching each student, rather than by separate supply-side considerations like content, timetabling and teaching interests. Next generation assessment should advance each student's experience, not least by using technology to leverage much richer forms of social engagement.

Assessment may not be validated against sound standards which assure quality. Much assessment is created as conservative adaptations of existing education and management practices, so framed in opaque ways with respect to legacy resources and with little subsequent validation. Assessments may be implemented differently across different contexts, giving rise to equivalency considerations which are negotiated rather than referenced to objective criteria. There may be only vague minimal requirements or specifications for common assessment practices. Next generation assessment should show fidelity to scientific standards rather than past practices.

Educational and practical decisions can be made without reference to financial considerations. While assessment is resource demanding, management focus remains on aggregate cost identification followed by reduction inclinations, rather than on building within a broader and potentially transformative financial context because full costing remains difficult. Financial models can focus on the aggregate cost of things which are easy to capture, and may not routinely take account of potentially significant resources associated with student time, exams, and online content development. At times, tasks may be deployed which are not marked/scored or factored/computed into students' overall outcomes. Next generation assessment should link education with financial models.

All up, there can be a lack of a coherent 'architecture' that brings together relevant facets of assessment. There are less commonalities in assessments than might optimally be the case given a more considered design. Assessment information and practices can be spread between different people, teams, systems and materials, which may be sustaining inefficiencies. Collaboration across units/fields may not be as extensive as it could be due to management structures or fidelity with legacy practices. Indeed, much assessment involves controlled and cautious translation rather than novel, excellent and efficient innovation. While it is core to education, there are often not yet people employed solely as assessment specialists who take overall expert responsibility for the quality and productivity of assessment. Substantial assessments such as exams and group work can be included by default, and without clear educational or commercial rationale. Next generation assessment should be framed by an educationally and practically cogent architecture.

Case Study 1 presents a fictitious case study application compiled from observation of various assessment practices. It details an 'anything goes' scenario, using the example of assessment in a particular subject. While it is core to education, people are not often employed solely as assessment specialists who take overall expert responsibility for the quality and productivity of assessment.

> Case study 1: Searching for better

A subject has been taught by the same academic for over a decade. With broad alignment to the overall degree, this teacher designs the teaching materials (overall plan, teaching resources, student readings) and teaching arrangements (mix of seminars and lectures, consultation hours, administrative supports). The teacher also produces the subject's assessments. While the teacher works largely alone, the assessment is framed by broader considerations such as degree structures, accreditation criteria, resource constraints, colleagues' expectations, and personal preferences.

The subject has many different assessments. Specific tasks include a three separate individual essays, a separate group assignment, and an end of term three-hour exam. Students are also encouraged to submit fortnightly reflective syntheses.

Viewed from the perspective of a well-designed assessment regime, there are problems. Total word count for each student is likely to gross more than 8,000 words. No feedback is given from the reflective syntheses or from the exam. The group assignment was inherited as it had no home elsewhere in the degree. The exam was included given norms and traditions across the faculty. The tasks do not interlink or accumulate, and are not synchronised to support a smooth learning experience. No tasks are aligned in any operational or technical respects with similar such tasks across the faculty. Detailed instructions vary, as do marking rubrics and grading schemes. The assessments are revised incrementally on an annual basis without obvious reference to other teaching practices or materials. The tasks are not built with reference to any broader technical standards. Little use is made of prior materials or analysis of prior experiences. The paper-based exam consists of a small number

of very open-ended questions which are marked by the teacher and one doctoral assistant in single day using a broad rubric with limited pre-calibration or cross-checking. The length of each of the essay tasks is given by assumed professional practice customs rather than the tasks to hand. There is little assurance that the tasks align with expected subject learning outcomes. While the university has invested heavily in sophisticated online platforms this subject makes little use of such functionality. While the teacher insists on managing the whole assessment process due to academic autonomy, this creates workload issues given that the work must then be juggled with reference to the teacher's unrelated research agenda. No consideration is given to finances beyond the teacher's workload and small direct salary expenses associated with the doctoral teaching assistant. Requirements which are somewhat arbitrary to the subject are given quarterly by a 'small working action group' of the faculty's education committee. Though there is not moderation across years, each year most students score in the mid-to-upper range of the grading distribution.

Overall, the experience is bumpy and perplexing for students, and they invest as much in trying to 'game' the process as in the sought learning itself. But students generally like the teacher so rarely complain and there seems little impetus for change.

It would not be this way if we started from scratch. Time to spring forward rather than look back. Together, these insights offer ideas and increased rationale for redesigning assessment. They affirm the particular significance of looking at what assessment is and what tasks are used, how these are deployed, and the contexts which shape deployment.

> Assessment Architecture

The designed assessment architecture consists of a family of models and processes. The models specify the educational and commercial characteristics of assessment. The processes detail the steps involved in assessment redesign. Together, the architecture provides a platform for high quality and efficient assessment. The architecture focuses at the task level rather than on broader assessment governance or structures, or on any narrower materials that make up specific tasks. For instance, it does not span broader social considerations relating to plagiarism, integrity and equity. These broader matters vary across delivery contexts, notwithstanding increasing standardisation against norms. Macro- and micro-level improvements can build on a progressive task-level architecture.

Flowing from the design perspectives and reconnaissance of existing practice above, the assessment architecture includes (Figure 7) an outcome model, resources model, experience model, standards model, and business model.

The outcome model defines the focus and nature of sought knowledge, practice and social outcomes, and links this with evidence captured by the tasks. The model structure is a simplification of existing mapping resources and approaches. It seeks greater specificity on the focus of the outcome, going beyond a general description. It also seeks information about the nature of the outcome to help understand its relevance and characteristics. It requires that tasks be aligned explicitly with outcomes, helping to ensure that minimally sufficient evidence/observations are available to demonstrate each outcome.

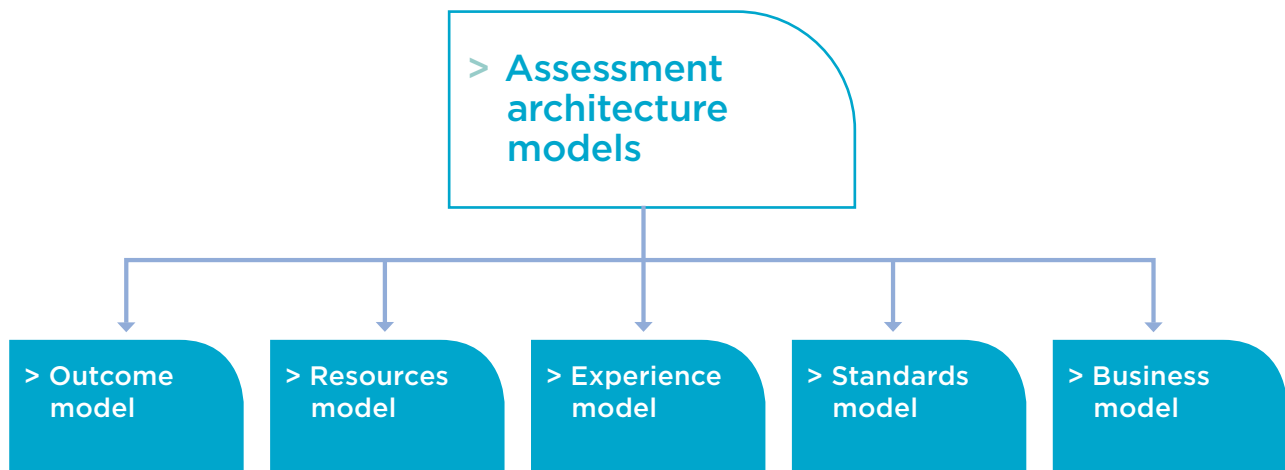


Figure 7: Assessment architecture models

“Improving the experience of assessments is essential to improving student engagement including retention, improving learning outcomes, and reducing staff time.”

The resources model specifies the actual tasks and other materials required to procure evidence from students. These materials include tasks and also supporting materials, rubrics and reports. Rather than defer to conventional/colloquial terms which are very loose, the resources model defines tasks in terms of their precise functional nature. The purpose of this is that it helps map the actual tasks in play and better control and specify these. There are efficiency and quality improvements to be found in more transparent practice, which, as with professional assessment work, must begin by clear specification and documentation of tasks.

The experience model incorporates a different perspective, focusing on what students are asked to do, hence their experience of the assessment. This helps ensure the assessment makes sense practically and in terms of how people experience the assessment. Improving the experience of assessments is essential to improving student engagement including retention, improving learning outcomes, and and reducing staff time.

The standards model details the technical, substantive and practical facets of assessment. They provide broad scientific, educational and management foundations for stepping beyond conventional forms of assessment practice. The standards have been sourced and distilled from a wide range of resources that, over the last half century, have defined the optimal and achievable properties of higher education assessment. All standards are applicable to all tasks. Such standards are necessarily normative ideals and it would be difficult, if not impossible, for any task to fully meet these expectations, but it is important, and distinctive, that critical consideration be given to each standard in task design, development, implementation and review. The quest is not that a task be flawless, but that it have known and optimised properties.

The business model focuses on the cost of assessment activities. It covers staff and material costs associated with development, including research, meetings, software development, consultation, and review activities.

It embraces costs associated with operations like managing student engagement, proctoring, marking and verification, and reporting results. Indirect costs associated with student time and materials are incorporated into the experience model.

It is important to clarify the distinctive nature of this architecture in higher education. In very rudimentary (and not uncommon) higher education it is typical for there to be specification of broad outcomes, formation of often very boilerplate resources, and consideration of the extent to which work is feasible within institutional/practical constraints. The assessment architecture takes a very different approach. Rather than see assessment as activity that derives from a broader teaching and curriculum or support process it instead enacts standards and approaches forged in education and measurement sciences as a basis for more principled development.

The architecture processes detail the steps involved in any redesign. The key processes (**Figure 8**) include setting foundations, analysing materials, confirming the redesign, and validating materials. Certain processes can be absorbed into and augment existing learning design procedures. Others are disruptive and require new infrastructure and perspectives.

By way of a brief overview of the generic processes:

- > Setting foundations involves building relationships with relevant personnel, and identifying rationales and approaches for development. These relationships are important because this new approach to assessment is necessarily disruptive of much prevailing practice.
- > Analysing the unit includes mapping the outcomes and evidence, taking stock of existing materials, and distilling observations to shape the redesign.
- > The heart of the redesign involves generating options, revising and specifying the resources, and auditing against standards.
- > Validating the materials includes developing tasks and associated rubrics, forming scoring protocols, and evaluating the materials.

The assessment architecture provides the platform for ensuring that higher education providers implement high-quality and efficient assessments. It consists of a family of general models that specify the educational and commercial characteristics of assessment. It consists of a series of generic processes that can be aligned with and augment existing design practice. Exactly how this assessment infrastructure is used depends on need.

Developing new tasks is the most comprehensive application of the architecture. It would involve stepping through each of the models and processes mentioned above. Each of the six integrated models is deployed throughout this process to both guide, document and refine the production of improved assessment.

Given that much assessment already exists, renovating existing tasks is likely the most common application of the architecture, and involves using it as an audit tool to diagnose risk and improvements. An initial scan might reveal the need to apply the model in some ways but not others, for instance to use the resources model to

parameterise and align existing tasks, or to use the outcome model to map materials to outcomes. A further broad scan might then help ensure the quality of materials.

Assuring tasks might simply involve application of the analysis and validation processes. This involves mapping tasks against the resources, outcomes and experience models, and then confirming properties against the business and standards models. Rather than a series of tasks, this audit would process a series of mappings and summary report.

Though fuelled by best intentions and masked by seemingly satisfactory outcomes, this briefing has asserted that much contemporary assessment is hardly optimised practically, technically or substantively. Application of the assessment architecture can remedy the situation. Ultimately, the architecture needs to yield better or great alternatives for even the least interested, able or time-poor academics. Case Study 2 constructs a picture of how such work can play out in practice.

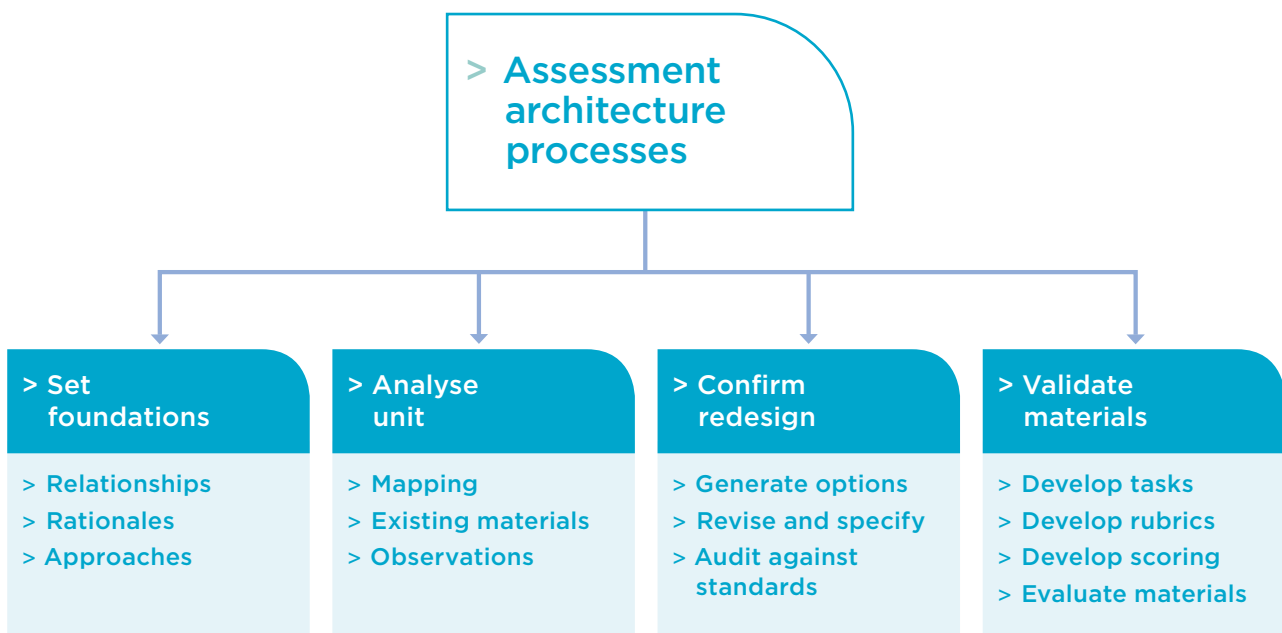


Figure 8: Assessment architecture processes

Though fuelled by best intentions and masked by seemingly satisfactory outcomes, this briefing has asserted that much contemporary assessment is hardly optimised practically, technically or substantively. Application of the assessment architecture can remedy the situation.

> Case study 2: A designed future

This briefing sketches a desirable long-game and it is useful to document what education professionals can do in coming months and years. Waiting a generation for evolution or revolution will not yield growth. Full-steam work is needed.

First, get assessment out of the closet and create or refresh conversations about what students know and can do. Big organisations swing towards big structures and tribes, but new approaches don't always align with existing roles. A Dean at one large university cut through by bringing together people doing different work on assessment for a roundtable meeting. People were working as managers, teachers, administrative support staff, academic assistants, learning designers, and learners. Drawing next-generation ideas into play, she created fresh dialogues about stagnating policies, new cross-cutting assessment collaborations, and three subject-level pilot initiatives to seed broader reform. Everyone had a chance to speak, and with differing views registered their interest in trying alternative approaches to assessment.

The Dean asked each of the pilot initiatives to plan how to build more productive forms of assessment. One group already had a well-estimated mapping of learning outcomes derived from earlier professional accreditation work but saw that they needed to further specify tasks and explore options for consolidation and alignment. For another subject only broader curriculum outcomes existed, and the team saw the need to distil more assessable learning outcomes then move from legacy essays and exams to a much more varied array of individually nuanced assessments. The third group had well-articulated outcomes mappings, task specifications and alignments, and deployments which made good use of technology to sustain each student's experience. What this group lacked was evidence of quality and efficiency.

After initial forays lasting a month, the Dean reconvened the teams to discuss their plans. Team leaders brought along other colleagues. Drawing on ideas aligned with those put forward in this briefing, the Dean created teams that could help with the development work ahead. People with experience in outcomes specification and task specification were asked to partner regardless of discipline or role with teachers who had yet to establish or fully articulate these foundations. The Dean invited people with rich experience using existing technologies to deploy more transparent forms of assessment to partner with the subjects needing task renovation. Finally, the Dean engaged assessment and finance experts from across the university to build templates for investigating quality and productivity.

The Dean asked for delivery of redesigned assessments in a year. She projected four milestones to guide design and development along the way—gather as much information as helpful, define redesign ideas, build prototypes, and test solutions to scale. Bi-monthly seminars were planned for the educators engaged to come together and share progress and problems. Further topic-relevant experts were called as consultants to advise about work along the way. An expert on academic work was asked to advise on new configurations and approaches to teaching. The university's Director of Learning Platforms gave insight into LMS plugins that could help give life to and scale the prototypes. Students were called into team meetings to test refined task formats and flows. Along the way, the Dean was careful to sustain energy on producing efficient and effective assessments rather than on the varying assessment expertise/roles of people contributing, the epistemological nuances of the subjects involved, or any hint that institutional policy was a pseudo barrier rather than enabler of progress.

At the end of the pilots, three redesigned next-generation assessments were produced with toolkit reports for use by others across the university. Particular effort was made to assure scientific, educational and management standards of the redesigned tasks, with the Dean seeing this as a way of building confidence in the integrity of the degree among students and graduate employers. By applying clear tests to the new infrastructure, the project was able to assure things that people ask about such as coverage and authenticity, practicality and responsiveness, and validity and interpretability. For a task in one subject, for instance, the teaching team was able to differentiate from rival programs by proving how investing in more in-depth teaching helped learners to master documented skills which employers had been seeking. The Dean evaluated the initiative and reported that despite the (largely sunk) start-up costs involved, she emerged with more engaged and efficient education infrastructure.

To lead this project the Dean drew on her deep expertise of higher education and student learning and assessment. Driven by a broad vision and through detailed steps she patched together the rudiments and platforms needed to create next-generation assessment. Not all leaders have time, relevant experience, or interest. The architecture articulated in this briefing seeks to make better assessment easier for all.

> Steps Ahead

Just because good assessment needs academic magic does not mean it should be immune from science. With a view to articulating next generation assessment this briefing has touched on just a few necessary and maybe sufficient conditions. There is so much to be transformed.

A few diagrams sketch out the kind of change required. Drawing on blue ocean logic, **Figure 9** presents value curves which caricature the three areas of assessment.¹ Through a huge generalisation, it differentiates these eras in terms of the five models of assessment architecture. Traditional practice may have been woven into an engaging educational experience but it lacked articulated (hence aligned) outcomes, parameterised tasks, scientific validation, or financial transparency. Expanded ('stretched') practice may introduce more alignment and more generic tasks, though at the expense of learner

experience and little fidelity to scientific or financial standards. By design, next generation assessment should be high on each of these dimensions.

Figure 10 charts the kind of collaborations among key assessment participants that seem likely and desirable to emerge as assessment transforms. The pictures show assessment shifting from something that single universities and single academics impose upon recipient students, through broader collaborations that involve interactions with various education services companies and industry agencies/networks, to much more sophisticated work which involves multilevel co-creation of planning, development, implementation, analysis and reporting. Clearly, there are many more individuals and interests involved, reinforcing the need for architectures which render such complexity workable in sophisticated ways.

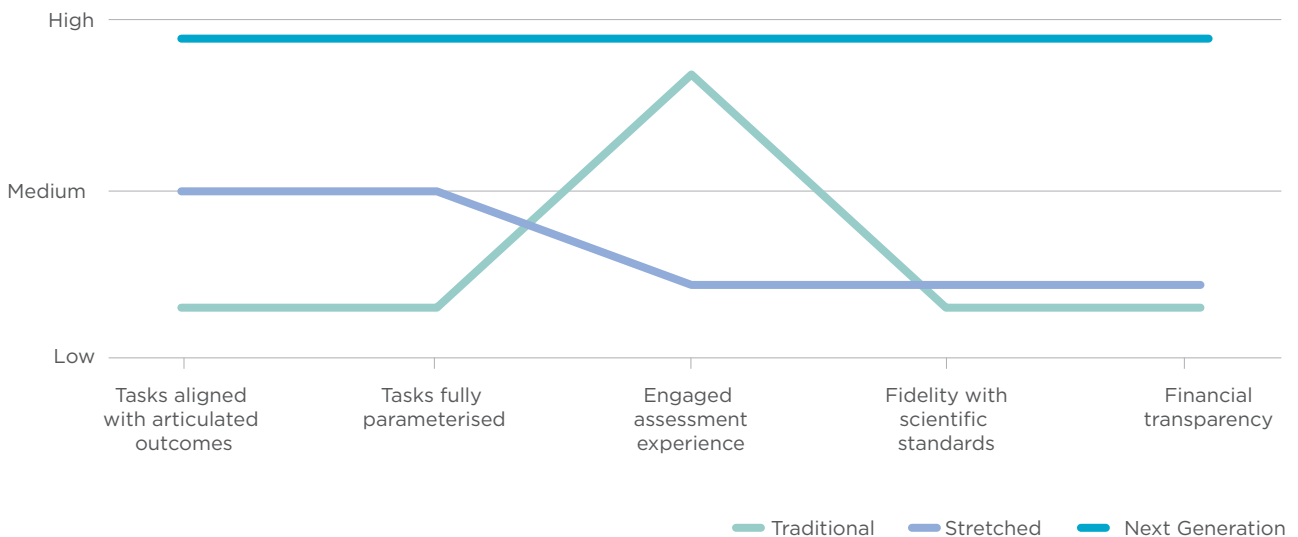


Figure 9: Value curves for three assessment eras

¹ | Kim, W.C. & Mauborgne, R. (2005). *Blue Ocean Strategy: How to create uncontested market space and make the competition irrelevant*. Boston: Harvard Business School Press.

> Traditional

Participants	Phases				
	Planning	Developing	Implementing	Analysing	Reporting
Government					
Industry					
Company					
University					
Academic					
Student					

> Stretched

Participants	Phases				
	Planning	Developing	Implementing	Analysing	Reporting
Government					
Industry					
Company					
University					
Academic					
Student					

> Next Generation

Participants	Phases				
	Planning	Developing	Implementing	Analysing	Reporting
Government					
Industry					
Company					
University					
Academic					
Student					

Figure 10: Next generation collaborations

“Surely the first development step involves mustering the courage needed to leap into uncertainty. Activating next generation assessment will need initial investment in building platforms that promises technical and financial returns.”

Figure 11 charts the nature and timeframe of innovation. It forecasts that traditional practices will decline in coming decades, partly as a result of demographic shifts. The fruits of signature collaborative developments pioneered in recent decades along with dividends of commercial activities and political projects will be distilled into optimised next generation practices. This innovation will flow across decades, relying as forecast in this briefing on a confluence of technological and organisational factors. More optimised practices will start to emerge.

There may be no quick or maybe cheap fix. Surely the first development step involves mustering the courage needed to leap into uncertainty. Activating next generation assessment will need initial investment in building platforms that promises technical and financial returns. The value of sharpening a business case for such change flows from the huge but as yet unreformed nature of higher education assessment. The odds of success seem good given assessment is one of the final multi-billion dollar parts of critical global infrastructure yet to be touched by the wand of productivity reform. Then comes the leadership to work through the distrust, subversion and denial that blocks change. Not everything is low hanging fruit. The rest is easy—coding minimum viable products, evaluating performance and integrity, negotiating legal peculiarities, building evidence of progress, and upskilling education engineers to lead future practice. These are the next steps, already underway.

It is important to qualify that computing technology is not the panacea for assessment. It may help people to eschew anachronistic practices and inspire reform. The architecture advanced here can be given easy software life, not just with simple MCQs and mechanical teaching functions, but with much richer forms of natural processing, emotional recognition, expressive interpretation, playful responsiveness and intuitive analytics. People will expect the kind of augmented and spontaneous interactivity that has infiltrated other facets of contemporary life, from banking to travel. They will seek instantaneous support when stuck, as with the collegial days of old. Employers, educators and learners will seek authenticated evidence of learning. Educators will need software that outwits the robots that bandits commission to cobble together fake assessment. Chief operating officers will need fresh proof that people are better than silicone at designing, proctoring, marking and feeding back. The world needs intelligent not stretched assessment. Computers give a technological fulcrum for lassoing a more productive future.

Next generation assessment will create much value. It will help learners engage with higher education. It will help faculty augment student learning and ensure the quality of education. It will help universities grow and improve. It will build novel capability and insights of global relevance. It parlays promise and underpins for the viability of future higher education.

	1990s	2000s	2010s	2020s	2030s	2040s
Traditional						
Collegial						
Commercial						
Political						
Optimal						

Figure 11: Confluence of innovation

100010111101011 IN EDUCATION, TECH-
NOLOGY CAN BE A.....
LIFE - CHANGER, A GAME CHANGER.

CAN CONNECT STUDENTS TO THEIR
PEERS IN OTHER PARTS OF THE WORLD.
IT CAN BRIDGE THE QUALITY
GAPS.

JORDAN 100010111101011 QUEEN RAINA OF

100010111101011 IN E
100010111101011



Online
Education
Services

Reimagining online learning

> **Call Us**

Australia

+61 (0)3 9956 0800

United Kingdom

+44 (0)20 3890 6910

> **Email Us**

enquiries@oeservices.com

> **Mailing Addresses**

Australia

Level 6, 600 St Kilda Rd
Melbourne, VIC 3004

United Kingdom

14 Gray's Inn Rd
London, WC1X 8HN