

Assessment Assurance in the AI Era – Final

James Dalziel, with Andrew Bain, May 2026¹

Executive Summary	2
Background.....	4
Higher Education Assessment First Principles.....	5
Problems with Assessment and Unit Scores	8
Two Lanes and Program-Level Assessment in the AI Era	10
Unit Level Assessment in the AI Era	12
Assessment Implementation Issues	13
Sample Templates for Assessment Redesign.....	16
Current Sectoral Responses to Assessment Assurance	19
References	21
AI Use for this Paper	22
Appendix 1: Transitional Issues for AUT in 2026 and 2027	23
Appendix 2: HDR and Long Projects	27
Appendix 3: Reflections on “Massive Disruption” Scenarios for AI	28
Appendix 4: Microsoft Copilot Analysis of V2 Paper	29
Appendix 5: Key Guidelines and Exemplars to be Developed	30
Appendix 6: Short Paper on Essays with Oral Validation.....	33
Appendix 7: Exams as Valid Measures.....	35

¹ This version of the paper (v.5.2) consolidates discussion from the February and May 2026 Teaching and Learning Committee (TLC) meetings, the May Academic Board (AB) meeting, and feedback received from college staff and others since February 2026. I’m grateful for all the helpful feedback provided both recently and in relation to earlier versions of the paper.

Executive Summary

Purpose and Scope

This paper addresses the medium-term implications (18–36 months) of generative AI on assessment assurance in higher education, with a particular focus on AUT coursework awards. It proposes a strategic reworking of institutional assessment practices to foster academic integrity and graduate readiness in the AI era. Implementation with students would be phased in over two years (2027–2028), commencing in Semester 1, 2027, and would apply to all academics at all Affiliated Colleges for teaching of AUT courses and units.

Key Assumptions

- Technological detection of AI-generated writing will be unreliable as a primary solution.
- Student misuse of AI is likely to be high, undermining traditional assessment integrity.
- Policy-based prohibitions will be insufficient; structural changes to assessment are required.
- Students themselves are concerned about AI's impact on their learning and critical thinking.
- TEQSA requires institutions to redesign assessment to address the problems of AI.

Strategic Challenges

- Misalignment between current assessment types and desired AUT graduate attributes.
- The need to balance AI-assisted (“Open”) and non-AI-affected (“Secure”) assessments.
- Risks of misleading unit assessment scores when combining Open and Secure assessments.

Recommendations

1. Adopt a Two-Lane Assessment Framework:

- Open Assessments: AI-assisted, formative, focused on learning.
- Secure Assessments: AI-free, summative, focused on assurance.

2. Redesign Unit-Level Assessment Structures:

- Two permitted approaches to unit-level assessments:
 - Use Secure assessments in a given unit where 40% (minimum) of the total assessments for the unit are Secure, and the overall unit can only be passed if the Secure assessment/s is/are passed as a “Hurdle” (regardless of total mark).
 - Use Secure assessments in each unit where 60% (minimum) of the total assessments for the unit are Secure (no Hurdle is required where at least 60% of the total assessments for the unit are Secure).
- Adopt oral assessments (e.g., Q&A, viva voce) and/or invigilated assessments.
- Allow essays and presentations that include a Secure “validation” Hurdle.
- Open assessments can help to prepare AUT graduates for future ministry using AI, noting the importance of AUT students and graduates using AI in a manner which is consistent with the Christian faith.

3. Adopt Unit Assessment Templates:

- Implement assessment templates for units with combinations of Open and Secure assessments appropriate to the unit content and AQF level.
- Consider the use of pass/fail assessments, and grading bands (only), in addition to scored assessments.

4. Enhance Academic and Student Support:

- Professional development for staff on oral and invigilated assessments.
- Clear guidance for students on effective and ethical AI use.

- Support students with anxiety about high stakes, time-pressured assessments.
5. **Plan for Transition (2026-2027):**
- Trial new assessment formats.
 - Prepare infrastructure for Secure online and face-to-face assessments.
 - Explore AI tools to better understand their capabilities for Christian ministry.
 - Mandatory application of a new approach to assessments for some units during 2027: Category A units, and at least two units per full-time academic. Full implementation in all units from Semester 1, 2028.

Appendices address transitional planning, HDR/Long Project considerations, reflections on existential AI risks, a list of anticipated guidelines and exemplars for implementation, details on essays with oral validation as a Secure type of assessment, and reflections on invigilated exams as a form of Secure assessment which remains valid for some units.

Background

Generative artificial intelligence (AI) will significantly affect all of higher education over the coming years, particularly assurance of assessment. There is much interest across the sector and beyond in the issues arising from AI and in responding well to the challenges and opportunities it brings, exemplified in our Australian context for example by the Castlereagh Statement (Castlereagh Summit, 2026). There are immediate challenges to manage during the current “transitional period” as student AI usage begins to have an impact but is not yet pervasive.² However, the purpose of this paper is to consider the medium term (18 - 36 months) when student use of AI is sufficiently widespread so as to require a fundamental reworking of institutional assessment practices.

This paper has four background assumptions:

1. Problems with technological detection of AI writing will render this approach flawed as the primary solution to assessment assurance (e.g., Alplusinfo, 2025).³
2. Student misuse of AI – where it is a clear breach of academic integrity requirements – will be high (40%+, see Q5.3, AlinHE.org, 2024).⁴
3. Policy-based prohibitions against the misuse of AI in relevant assessments (and related student-focussed “communication” approaches) will be increasingly ineffective. Addressing AI misuse will require structural changes to the nature and mechanics of assessment (Corbin, Dawson & Liu, 2025).⁵
4. Students themselves are concerned about how excessive use of AI is reducing their abilities for critical thinking and sustained engagement with complex ideas – they recognise that AI is part of their future, but they are worried that excessive use of AI will be detrimental to their learning (e.g., Sulkin, 2025).

Similar concerns are evident in TEQSA’s September 2025 publication “*Enacting assessment reform in a time of artificial intelligence*” (Lodge et al, 2025). TEQSA notes:

Under many traditional assessment practices, the inappropriate use of gen AI compromises the assurance of and challenges compliance with Threshold Standard 1.4.4, which requires the demonstration of specified learning outcomes. Since detecting gen AI use with certainty in assessments is, at this point, all but impossible, we need alternative approaches to complement academic integrity processes. These approaches must either:

- permit gen AI use within defined parameters (recognising that allowing any gen AI use effectively permits its use for entire tasks in most instances)
- design assessments where gen AI use is irrelevant to the demonstration of learning outcomes
- restrict unauthorised gen AI use through direct supervision of learning and/or performance.

...Assessment and learning experiences should equip students to participate ethically, critically, and actively in a society where gen AI is ubiquitous.

² Transitional issues for AUT during 2026-2027 are considered in Appendix 1.

³ Apart from general detection problems, there is a concern that AI detection systems may disproportionately affect students who are non-native English speakers and some neurodivergent students.

⁴ While my presumption is that most students would rather not misuse AI, and this is supported by Diane Hockridge’s recent research on AUT student attitudes to AI, I believe there is insufficient appreciation of the vicious spiral that arises from some students misusing AI and getting the benefit of doing so, which is seen as unfair by other students, which leads to a spiral of growing misuse of AI by students so they aren’t disadvantaged relative to their peers.

⁵ A development which highlights how far AI’s capabilities have already progressed, in terms of the possibilities for misuse of AI in student assessments, is the emergence of Agentic AI. This technology enables AI agents pretending to be human students in Learning Management Systems, completing a range of unit tasks such as producing and submitting online forum posts (and responses to other posts) or completing Moodle quizzes, to make it appear as though they are an active “real” student. For an introduction to Agentic AI and its use by students, see this article by [Wagner \(2026\)](#).

...Rather than investing primarily in detection mechanisms, institutions need to emphasise the redesign of assessment to capture authentic demonstrations of student capability and comprehension.

TEQSA has further stated that it is now shifting from its previous “educative” phase about AI to an “enforcement” phase, where it expects all institutions to address the problems raised by AI. At the same time, TEQSA and others (Corbin et al, 2025) have noted that AI in higher education has the characteristics of a “wicked problem”, in that the problem is difficult to define, challenging to solve, interconnected with other problems, and lacks clear solutions.

The focus of this paper is primarily coursework degrees (rather than higher degrees by research – HDR) in fields with a high reliance on take-home assignments (such as essays) and other forms of written assessment tasks, such as the humanities and social sciences.⁶ Initial consideration of HDR concerns, and related issues for Long Projects, are provided in Appendix 2. Some broader reflections on the impact of AI on society, especially “massive disruption” scenarios, are provided in Appendix 3.

In keeping with the Christian Foundation of the Australian University of Theology (AUT), the ultimate goal of this paper is to understand how best to prepare AUT students to serve God’s church and God’s world in the coming AI era, and how AUT staff can best contribute to the education of these students, to the glory of God.⁷

Higher Education Assessment First Principles

Before considering the impact of AI, it is helpful to go back to first principles about higher education and assessment. When a higher education provider awards a degree, it is indicating that the graduating student has met the expectations of the relevant course. These expectations are typically described in terms of graduate attributes and course learning outcomes, which are supported by the unit teaching and learning activities, and demonstrated by results in assessments. The award of a degree is effectively a statement that the graduate has achieved a certain kind of knowledge and skills, and can perform certain kinds of tasks, and, in some contexts (such as theology), has developed a certain kind of character.

The link between assessments and the attributes of graduates is often called “constructive alignment” (Biggs, Tang & Kennedy, 2022, especially pp. xxii-xxiv). This is the extent to which the assessments provide evidence that the graduate has achieved the desired knowledge, skills and character attributes of the degree. For example, if a key attribute of theology graduates⁸ is the ability to talk about Christian ideas publicly (e.g., give a sermon), but graduates have only ever been assessed via written essays, the graduates may have demonstrated the knowledge component of the course (in the essay), but whether they have the skill of public speaking about that knowledge is untested, and hence unknown.

Some assessment is “formative” in the sense that the assessment process itself seeks to foster learning (sometimes described as “assessment for learning”). But for course completion and graduation, the key type of assessment is “summative”, which involves evaluating the learning that students have achieved (sometimes described as “assessment of learning”). While summative assessment is not necessarily based on scores⁹ (a “pass/fail” or “satisfactory/unsatisfactory”

⁶ For certain areas of software, engineering, architecture and some other technical and scientific fields, AI may become so embedded in routine tasks that the ideas presented in this paper may not apply. The focus of this paper is not only theology but a wide range of courses in the humanities and social sciences, as well as some professional disciplines (e.g. education, law).

⁷ While it is beyond the scope of this paper, AUT academics should continue to think critically about technology, especially AI, and its positive and negative implications for faith and society. The focus here is the more limited issue of how higher education assessment should be redesigned in an era where AI is pervasive.

⁸ “Theology” here should be interpreted to refer to Christian ministry broadly, rather than the specific subfields of systematic, historical and/or biblical theology, and hence it refers to all graduates of the AUT.

⁹ Also called “marks”, although I will use “scores” here to avoid confusion with text comments that assessors make at various points on assessments like essays, which are also sometimes called “marks”.

assessment can be a summative assessment), the most common form of summative assessment involves a task which is given a score, and students who fail to meet the required pass score for the collected tasks of a unit would fail the unit (and hence not progress towards course completion and eventual graduation). Even if a unit is only given an overall grade such as “Credit” (but not a score), final grades are often determined by combining assessment scores across relevant tasks.

Recent approaches to higher education encourage greater flexibility in assessment than many academics in theology realise. Current regulatory structures envisaged by TEQSA and the Australian Qualifications Framework (AQF) can support great variety in assessment types and marking approaches. For example, the use of pass/fail (or equivalent) assessment is being widely adopted in higher education, even in “high stakes” professions such as medicine (with a recognition that pass/fail does not preclude setting a very high bar for passing). A related option is to use grade bands (only) without a specific score – this allows for some level of differentiation in performance where the greater precision implied by scores is not relevant or warranted. There has also been a recent shift away from a heavy emphasis on numerical scores and “grade point averages” (GPAs) and towards assessment of competence, and mastery of the skills needed by a typical graduate (and assessment may focus on descriptive statements of ability rather than just scores).

This relates to a wider education trend away from “norm-referenced” assessment (“grading on a curve”) towards criterion-referenced assessment, where the focus is on whether each student has achieved the desired level of knowledge/skills/character for themselves, regardless of how well or poorly other students have performed. Under criterion-reference approaches, there is no minimum number of students who should pass (or fail) a unit – passing is simply a function of whether each individual student has achieved the minimum requirements for successful completion of the unit.

These assessment shifts are driven by a greater emphasis on the role of higher education as shaping students to be a certain kind of person for a certain kind of future life and work, rather than emphasising ranking students from the “smartest” to the “less smart”. This is most clearly seen when the focus of assessment is not ranking or scoring, but determining whether a student has achieved relevant learning outcomes by gaining the skills and knowledge to be effective in work and life as a graduate.

For theological education, this means graduates should develop appropriate skills and knowledge for the kind of Christian work they would be expected to do after graduating with their theology degree. While this varies from student to student, and degree to degree, the AUT has two broad categories of students – those training for vocational ministry, such as working as a church leader or missionary, and those training to gain greater Christian knowledge and skills so they can be more effective in a wider range of “lay” Christian roles, such as Bible study leaders, Christian community contributors, charity workers, effective Christians in secular workplaces, etc. In terms of length of study, those training for vocational ministry often engage in longer periods of study (e.g., 3 year Bachelor of Theology, Bachelor of Ministry, Master of Divinity), while those studying for “lay” roles often (but not always) engage in shorter periods of study (e.g., 6 month Undergraduate or Graduate Certificates, 1 year Diploma or Graduate Diploma).

While there are differences in expectations for these two broad categories of students, there are many commonalities that should inform a description of the attributes of AUT graduates – for example, they should:

- be effective communicators of the Christian message (e.g., in conversation or a sermon),
- have pastoral skills,
- have a deep knowledge of the Bible (potentially including knowledge of original languages) and the ability to apply this knowledge to the practical challenges of daily life,
- be able to think on their feet about giving advice and solving problems in ministry and life, &
- have a grasp of the theology and history of the Church that helps them to be reflective about the challenges of modern society.

To the extent that formation of Christian character is a focus of their degrees, this would also be an attribute of AUT graduates.¹⁰

AUT currently has four broad graduate attributes: students are to develop as Christian people, Christian scholars, Christian professionals and Christian leaders. Depending on the type of degree and length of study, the amount and kind of development in each of these four areas will vary, but at a high level of abstraction, there is much in common across AUT degrees regarding these graduate attributes. Nonetheless, a review of AUT graduate attributes is worth considering at this time, including whether ethical and effective use of technology (including AI) needs to be incorporated.

The problem is that, even before AI arose to challenge current assessment practices, there was often a mismatch between the types of assessments used in AUT units and the kind of knowledge, skills and character that is desirable for an AUT graduate – assessing preaching with an essay is an obvious example. This is the “constructive alignment” problem noted earlier. Regardless of AI, there are steps that AUT could take to use the flexibility available in modern higher education assessment to better align its choice of assessments with the desirable attributes of AUT graduates.

In practice, this means that while essays remain important as a way of testing Christian understanding and sustained argument, they would play a less significant role than at present (relative to other assessments that align with desirable graduate skills). In the place of some essays would be more assessment of the skills needed for effective Christian workers – skills like preaching, pastoral care, the formation of Christian character, and thinking “on the spot” about challenging real world ministry scenarios. These kinds of skills need different kinds of assessments, including group discussion, individual “Q&A”, assessment of practical skills like preaching, and guided spiritual formation exercises. And many of these assessments may not need scores or ranking – they might be better assessed in terms of competency and mastery of skills (which could be “pass/fail”), and development of character and wisdom.

However, any change to AUT assessment should not lead to lower academic standards, or a plethora of undemanding “feel good” assessment exercises. The AUT has always maintained high academic standards, and should continue to do so. The challenge is to maintain high academic standards while implementing assessments that evaluate whether or not our graduates have developed the knowledge, skills and character that are needed for a variety of roles in Christian ministry, and as Christians in modern society.

As for the traditional long-form, take-home essay, the role of this assessment tool will unavoidably be changed by the impact of the AI era. Essays will still be used, but the likelihood of AI use in essays will change their significance in the overall “toolkit” of assessment methods. One approach to retaining a significant role for long-form take-home essays would be to incorporate an oral assessment of the content of the essay. Details of how this approach would be conducted, and in what circumstances, will need further consideration.

The issues discussed above require rethinking the kinds of assessments that are used in AUT units, as well as the way that marking processes are conducted, especially the role of scores on assessment

¹⁰ Across the AUT Consortium, some Affiliated Colleges seek to achieve Christian character formation within AUT courses, such as through Guided Spiritual Formation units, while other Affiliated Colleges (and/or their denominations/stakeholders) seek to develop Christian character mostly outside AUT courses. Formation could be the focus of separate denominational/missionary agency training initiatives, or it could be the focus of a local Affiliated College certificate which is not part of the AUT’s formally accredited courses. While some degree of Christian formation would be anticipated among participants in all AUT units, a key issue for the current context is whether this formation is assessed, and if so, is there a realistic prospect that a student could “fail” Christian spiritual formation (and if so, what does failure mean in practice, and how might a student “pass” in the future?). Formation is an issue that deserves wider consideration outside the context of the current paper, but it is worth noting that while explicit Christian character formation is an option within AUT courses, it is not a primary focus of how some Affiliated Colleges use AUT courses, and hence AUT course structures needs to be flexible about the role of formation in its graduate attributes.

items and how these are combined for overall unit scores. The next section considers problems with scoring, and how the rise of AI intensifies these problems.

Problems with Assessment and Unit Scores

Adding up scores across assessment tasks is not without problems (Dalziel, 1998), but in the context of AI there are new concerns. Consider the simplified case of a unit with two equally weighted assessment tasks, a take home essay and an invigilated assessment¹¹ essay. In a typical scenario, a student who scores 50 on each task would receive a final score of 50 and pass the unit.¹²

The overall unit passing score is an indicator that, given two different assessment types (a long timeframe essay task supported by easy access to supporting resources versus a short timeframe essay in a constrained environment without supporting resources), the student has performed sufficiently well in both assessments to be considered to have met the minimum requirements of the unit, and hence passed. The topics of the two essays would typically be different, indicating sufficient breadth of understanding of the material covered in the unit. While the match between individual assessment performance and the full range of topics covered in a unit will almost always be imperfect, educators effectively set a “good enough” level of performance when they adopt a passing score for different kinds of assessment types and topics in order to pass a unit as a whole.

How should educators react to a case where a student gets 40 on one essay and 60 on another essay? Given that the final unit score is determined by averaging the two marks (leading to a score of 50), the overall passing score is saying, in effect, that the lower-than-passing-level-performance on one essay is offset by the above-passing-level-performance on the second task.

Now consider a case study where AI is available to students (but forbidden for use) for take-home essays but AI is not available for invigilated assessments. If a student gets a score of 75 for the take home essay, and 25 for the invigilated essay, the final score would be 50, and the student would pass. But stepping back from the scores, has the student demonstrated the minimum “good enough” level of understanding that is required to pass this unit in order to be a competent graduate?

If the societal context is one where graduates are assumed to always have access to AI in their lives, and will never experience tasks that require a short timeframe response without supporting resources (including AI), then a case could be made for the student passing this unit. But this would mean a complete reworking of graduate attributes and learning outcomes for a world where people rely on AI at all times for all tasks. And even if that day is coming, we are far from it now.¹³

So the appropriate answer for the current context is that, even where AI is assumed to be a part of the future lives of graduates, they will still encounter many situations where they will need to use their knowledge and skills in time-limited contexts where they lack easy access to resources (such as AI). For example, if one of the future tasks of a theology graduate is to provide comfort to a

¹¹ For the purpose of this paper, an invigilated assessment is assumed to be a context where students would not have access to AI, either as a written assessment or using locked-down computers that are unable to access AI. Invigilated assessments may be written exams, but there are other kinds of assessments that can be invigilated. It is worth noting that for students not on campus, remote exam software designed to lock down computers may not be sufficiently robust to ensure AI is not available (such as via a second device, an earpiece, etc) unless an invigilator is physically present. Details for effective invigilation in the AI era will need further development. Cases where an invigilated assessment deliberately includes the use of AI would require special guidelines.

¹² More complicated structures can involve weighting of two or more assessment tasks, with the final score arising from the weighted combination of scores according to a preset formula.

¹³ We also have very little idea what the formation of character will look like in a pervasive AI world.

distraught person during (real-time) personal interaction, the graduate needs to be able to do this “on the spot” without relying on external resources.¹⁴

Regarding the two-assessment case study, a student who gets 75 on a take home essay but 25 on an invigilated essay shows that without access to AI, his/her understanding of the unit is very poor, and hence the student should not pass and proceed towards graduation. That is, the higher education provider does not have sufficient assurance that the student is meeting course requirements based on assessment outcomes. How can this assessment assurance challenge be addressed?

One option would be to only use assessments that were not “AI affected”,¹⁵ such as two invigilated essay assessments. A related option would be to adopt a variety of assessments types so long as all of them are unaffected by AI, for example, an oral assessment (e.g., “viva voce”) and an invigilated essay assessment. Unfortunately, some of the current suggestions for assessment types to avoid potential AI misuse are no longer effective as AI becomes more advanced.¹⁶

However, this approach poses a different problem – if AI will be widespread in the future and affect much of the working lives of graduates, then competence in using AI effectively and ethically will be a relevant graduate capability in its own right.¹⁷ This is true not only in terms of AI’s general impact on society and the church, but also in the specific ways AI can assist Christian ministry, such as assisting with sermon preparation, translation, and summarising feedback (for an example of potential AI use in Christian ministry, see Max7, 2025). There will also be specific ways AI contributes to Biblical research, such as by identifying subtle textual patterns in textual criticism. For a discussion of how AI could contribute to Biblical studies, see McGrath, (2025).

As Bearman et al (2024) note, one of the most important skills for students to develop in the AI era is evaluative judgement, particularly about AI outputs.¹⁸ This suggests that relevant AUT assessments need to be designed to incorporate the use of AI in a constructive way, while not undermining other

¹⁴ And even in a futuristic world where a person offering comfort has access to “in ear” real-time AI advice (or text advice seen through AI-enhanced “smart” glasses), the comforter would still need to be able to critically evaluate the suggestions of the AI to determine whether these are helpful or not.

¹⁵ By “AI affected” I mean assessment tasks where AI is available to a student when doing the assessment. Student preparation for an assessment may involve AI, but for typical invigilated assessments, the student can’t use AI during the assessment itself. To avoid undue impact of AI-based preparation, invigilated assessment topics must be sufficiently unpredictable, so that prior memorisation of AI answers isn’t a viable strategy.

¹⁶ For example, some have assumed that diaries or other personal reflection exercises could avoid AI misuse – however, AI is particularly good at “faking” personal-sounding reflection if given relevant instructions. For some examples of the capabilities of recent AI tools, see [this TEQSA video from Danny Liu](#). AI can also create compelling discussion forum postings (see [this TEQSA video from Ben Miller](#)), even drawing on comments from previous students in the forum in a response. For assessments where students show their developing ideas over time (so that the marker can see the evolution of student thinking behind a completed essay), AI will no doubt prove adept at taking a “completed” essay (which could be AI generated), and then “inventing” a history of progressive development of these ideas over time (including “wrong” ideas which are subsequently discarded!). Finally, AI audio and video generation is reaching the stage that it will be able to fake a presentation by a student (using the student’s own voice and appearance) based on AI generated content – so that despite the appearance of the student talking about their ideas, both the content and the video were created by the AI with minimal student input. This AI capability is why “live” interactive student assessments are required. For an example of this problem see [this second TEQSA video from Danny Liu](#).

¹⁷ One of the significant benefits of AI for higher education is the ability of AI chatbots to respond to individual student questions about topics, and to customise answers based on prior student understanding, and persistent misunderstanding. There is significant scope for carefully selected and configured AI chatbots to assist academics by providing personalised feedback to students in parallel with academic-led activities. There will also be other non-teaching and learning benefits to higher education from adopting AI chatbots, such as 24x7 advice for students on enrolment and course progression requirements and related administrative issues.

¹⁸ Additionally, students need to learn to avoid unhelpful habits such as excessive or poorly considered cognitive outsourcing to AI programs. See Lodge and Loble (2026).

attributes of a competent graduate¹⁹. Put another way, what is the ideal balance of “AI-assisted” and “non-AI-affected” assessments that lead to assurance of graduate attributes?

Two Lanes and Program-Level Assessment in the AI Era

One approach is to adopt two “lanes” for assessment (AI-assisted or “Open” assessments versus non-AI-affected or “Secure” assessments, see Bridgeman & Liu 2024a). AI-assisted (Open) assessment can be considered “assessment for learning” and non-AI-affected (Secure) assessments can be “assessment of learning”. Each unit assessment task is considered in terms of how it contributes to graduate attributes, keeping in mind the need for assurance of assessment.

An “Open” assessment is not necessarily “an assessment in which the use of AI is a requirement of the assessment instructions”. Rather, it is simply a recognition that it is not possible to effectively restrict AI usage for many kinds of traditional assessments, such as take-home essays, and hence markers should presume that at least some (possibly many) students will use AI in these circumstances (even when they had been told not to do so). “AI-assisted” is not the same as “AI-endorsed”. Nonetheless, one important type of Open assessment in the AI era is where the assessment task explicitly expects students to use AI, and provides advice and guidance on appropriate and ethical AI use which is consistent with the Christian faith,²⁰ in order that graduates are prepared for a future where AI use will likely be pervasive, including in Christian ministry. In the context of the AUT, this can involve considering how AI can contribute positively to future ministry tasks (as illustrated by the examples in Max7, 2025), and how it can be used in a manner aligned with Christian ethics.

One implementation of the two lanes approach is to consider the differential use of these lanes across the course-level structure of the degree. This can be called “program-level” or “programmatic” assessment, in that the assessment structures operate primarily at the level of the whole program of the degree, not simply at the unit level (Bridgeman & Liu, 2024b). This approach is described as “Pathway 1” in TEQSA’s 2025 “Enacting assessment reform in a time of artificial intelligence” (Lodge et al, 2025).

An example of programmatic assessment is in medical degrees, where there are many smaller scale (often formative) assessments, especially hands-on clinical tasks, but the key summative assessments are infrequent “barrier” exams (e.g., once a year), where students are tested on all course material at once, and students who do not meet the minimum requirements for progression (based on the required graduate attributes of a competent doctor) are unable to continue into future years (that is, the exams are a genuine barrier to progression).

Certain types of degrees are more or less suitable for program-level assessment. For example, while it can work well in medicine, it is more difficult to implement in generalist degrees which involve flexible unit structures and a significant variety of potential units (especially with large numbers of elective units) – such as at the AUT. As each unit has its own assessment structure, and the variety of unit combinations available in generalist degrees is extremely large, it can be nearly impossible to assure that a particular set of units will meet programmatic assessment goals if the assessments in individual units don’t match the programmatic requirements.²¹ This is a key problem TEQSA’s notes with “Pathway 1” type solutions. For example, a generalist degree may contain units with barrier

¹⁹ More detail on integrating effective use of AI into AUT courses and assessments will be needed in the future, both for academics and students. Among the issues to be addressed are concerns about the ethics of using AI (such as copyright and power usage). There is also a potential equity issue related to all students having access to a sufficiently capable AI system – this may require AUT to obtain a site-license for an appropriate AI that is made available without charge to all students (e.g., Microsoft Copilot as part of a freely available suite of Microsoft Office tools).

²⁰ Noting that exceptions may need to be made for those who conscientiously object to AI usage.

²¹ A visual illustration of this problem is provided on slides 23 to 26 of my [presentation at the November 2025 AUT Professional Development Conference](#).

assessments, but if unit choice flexibility allows students to choose units without these barrier assessments (or without enough of them), the student can graduate without meeting program-level assessment requirements. This problem becomes even greater with shorter length flexible courses, such as generalist Undergraduate and Graduate Certificates and Diplomas.

Consider a four unit Graduate Certificate with three assessments in each unit (similar to example 6 in Bridgeman & Liu, 2024b). What should be the frequency of Secure assessments across this structure? One approach would be to have a “barrier” Secure assessment at the end of the second and fourth units, with AI use (Open) expected for the remaining ten assessments, including all assessments in the first and third units.

While there might be specific contexts where this approach could meet graduate attributes, there are several potential problems:

1. What does it mean for a student to pass a unit when no assessment in the unit has been assured against AI usage (that is, there are no Secure assessments)? If students use AI for every assessment, what have they learned and demonstrated apart from using AI?
2. Even in units with a Secure assessment, can high scores on Open assessments overcome a poor score in the Secure assessment? For example, if a student scores 75 on both Open assessments and 25 on the Secure assessment, the overall unit score will be $(75+75+25)/3 = 58.3\%$, which means the student will pass despite scoring poorly on the Secure assessment.
3. Does the course structure actually implement the “barrier” in practice? That is, if students fail the second unit (which has a barrier assessment), are they unable to study the third or fourth unit until they pass the second unit? What happens if units are being studied concurrently (e.g., units 2 and 3)?
4. If a student fails a unit and is unable to proceed, and the student appeals against this process as being unfair or otherwise detrimental to the student’s future prospects or mental health, does the institution have a sufficiently robust appeal review process to assure that students are not able to “sidestep” the barrier via appeals, and ultimately graduate without demonstrating the relevant graduate attributes?

Taken together, I believe these four problems mean that, apart from special contexts (like medicine), program-level assessment will not be an effective solution to addressing the hard problem of assessment assurance in the AI era for highly flexible course structures like those at AUT. This brings us back to unit-level assessment structures. This kind of approach is described as “Pathway 2” in TEQSA’s 2025 paper, and hence AUT seeks to implement a Pathway 2 approach.²²

The problems identified above are just some examples of how AI is a “wicked problem” for higher education (Corbin et al, 2025). As Corbin and colleagues note *“In the context of higher education assessment, the challenge posed by GenAI is better understood not as a problem to be solved but as a condition to be navigated.”* Any proposed solution will be imperfect, and will require ongoing refinement and iteration. Corbin and colleagues also argue against one size fits all solutions, and recommend *“permission to compromise”*, but they also note *“while it is true that wicked problems do not have correct solutions, they do have better and worse responses”*.

I can see two main options for implementing a Pathway 2 approach that directly address the assessment assurance problem at the unit level. Given the clear consensus of the May TLC meeting on this matter, and in light of the flexibility needed to address AI as a “wicked problem”, it is recommended that we adopt two permitted alternatives for unit assessments, which reflect these two main options:

1) Secure assessment totalling at least 40% of the total unit marks which incorporate a “Hurdle” requirement;

²² “Pathway 3” in TEQSA’s paper is a hybrid of Pathway 1 and 2, but given the unresolved problems with Pathway 1 identified here, Pathway 2 is the preferable approach for AUT.

2) Secure assessment totalling at least 60% of the total unit marks, which do not need to incorporate a “Hurdle” requirement.

Unit Level Assessment in the AI Era

It is important to note that the specific arrangement of Secure assessment(s) in a unit remains a decision for academics. While an invigilated assessment is likely to be widely adopted for practical reasons, it is not a mandatory requirement for every unit, as long as other appropriate Secure assessments are used (for example, a unit could have a Secure oral assessment and no invigilated assessment, together with an Open assessment). In the same way, Secure oral assessments are also likely to be widely used, but they are not mandatory (for example, a unit could have an invigilated assessment, together with an Open assessment). And for completeness, some form of Open assessment will also be widely adopted, but it is not mandatory (in both of the preceding examples, the Open assessment could potentially be removed, as long as relevant course/unit learning outcomes are still achieved).

The first option requires each student to always pass a Secure assessment regardless of their scores on Open assessments. Returning to the earlier two assessment case study with a score of 75 for the Open take-home essay and 25 for the Secure exam essay, the student must now get at least 50 on the invigilated essay regardless of how high his/her score is on the AI assisted assessment. In TEQSA’s Pathway 2 approach, this is called a “Hurdle” requirement. The final unit score is no longer a weighted addition of scores, rather, it is a more complicated structure that ensures any Secure assessment must always be passed for the whole unit to be passed.

An example of the first option could be a unit with two assessments (both worth 50%): an Open take home assignment, and a Secure invigilated assessment. In this case, regardless of the score on the Open assignment, the student must score at least 50% on the Secure invigilated assessment in order to pass the unit (regardless of what their weighted average mark would be otherwise). For example, a student who scored 82 on the Open assignment, but 48 on the Secure assessment would fail the unit (because the student failed the Secure assessment), even though the weighted average mark would otherwise have been 65.

One of the challenges of this option is the potential for student confusion/concern about how their final mark is determined. Students may feel it is unfair that their overall weighted unit score was at or above the pass score, but they will fail the unit regardless. However, the potential for student confusion/concern will need to be overcome with discussion about the rationale for this structure, and clear indications of marking processes for each unit.

The second option seeks to achieve a similar outcome but without the explicit rule about unit failure if any Secure assessments are failed.

An example of the second option would be a unit with three assessments: an Open take-home assignment, a Secure interactive oral assessment, and a Secure invigilated assessment. In this structure, each assessment could have the same weight, and hence two thirds of the scores (over 60% of the unit total) arise from Secure assessments, while also incorporating an Open assessment to foster skills in using AI effectively (subject to appropriate student support for the use of AI). A weighted addition of scores in this case would not remove the impact of AI on unit scores, but it reduces it to no more than 40% of the total, and given that all students would be given access to AI for the Open assessment, scores on this assessment do not differentially impact overall scores in a way that advantages some students over others (i.e., compared to take current home assignments where, despite AI prohibitions, some students use AI and some don’t).²³

²³ It is also possible that the range of typical scores on AI-enabled Open assessments might be more limited than other assessments – that is, typical scores might cluster around the 60-70 range when AI is available,

There could be options to address potential problems arising from inflated overall unit scores due to the potentially high scores on the Open assessment, for example:

1. The scoring rubric for the Open assessment could be adjusted so that high scores were harder to obtain (scores are effectively “scaled down” via more challenging criteria).
2. The Open assessment could be graded as “satisfactory/unsatisfactory” only, not scored.
3. Greater relative weight could be given to the Secure assessment in calculating the overall unit score (e.g., the Open assessment could count for only 20% of the total score, while the other two Secure assessments could count for 40% each, rather than each assessment counting for one third).

For 1, a difficulty is that students who don’t use AI (including AI conscientious objectors) may be potentially at a relatively disadvantage in their results due to the proposed new harder marking. Unless a resolution to this problem can be found, 1 may not be practical.

Following discussion with Academic Board and Teaching and Learning Committee in late 2025 and early 2026 and then further discussion at the May 2026 TLC meeting, it is recommended to adopt both the first and second options discussed above as permissible approaches to Secure unit assessment – that is, each unit will have (at least) one Secure assessment, and in cases where the Secure assessments are worth less than 60% in total the Secure assessment/s will itself be a Hurdle (e.g., invigilated assessment, interactive oral assessment) or incorporates a component which acts as the Hurdle (e.g., and essay with validation exercise, where the validation exercise is the Hurdle). The key role of any Hurdle is that students must pass the Hurdle to pass the unit overall. The 40% minimum Secure assessment (with Hurdle) requirement per unit can be either a single 40%+ assessment item, or several smaller secure items which together comprise at least 40% of the unit mark. For example, it could comprise a single invigilated exam worth 50% or single larger oral assessment item worth 40%, or alternatively three relatively shorter secure invigilated tests worth 15% or 20% each or two relatively smaller secure oral assessments worth 20-25% each. Alternatively, a Hurdle will not be required if the total Secure assessments in the unit are worth 60% or higher (this requirement replaces the earlier “two thirds” option).

Assessment Implementation Issues

While an earlier version of this paper proposed full implementation of the requirements here from semester 1, 2027, further feedback has indicated that a staged approach over two years (2027 and 2028) is a more realistic implementation goal (given the workload involved and existing demands on staff). In this version of the paper (V5), the recommended approach is that all Category A units implement the required approach in 2027, with all other units adopting the approach across 2027 and 2028 subject to academic workload. Given a full-time academic may have up to eight units to change, it is proposed that each full-time academic be expected to implement the new approach in at least two different units in 2027, so as to avoid a large number of changes needed in their other units during 2028. A special exception would apply where a staff member is a unit coordinator for more than two Category A units in 2027: to ensure that such staff do not need to implement the new approach in more units than they can practically manage during 2027, any such staff would not be required to apply the new assessment requirements in more than two of their Category A units in 2027 (but could still apply the new approach in more than two units in 2027 if they wish to do so). Advice for students on the nature of the staged implementation will also be needed.

whereas Secure assessments might have a wider range of typical scores, say 40-70. The point of this observation is that due to the different ranges of Open versus Secure scores, Secure assessments will make a relatively greater contribution to overall differences among students for overall unit scores.

Regarding oral assessments, advice will be needed for both academics and students on how to best engage with this assessment approach²⁴. In the context of the concerns in this paper, this assessment needs to be conducted in a way that genuinely tests student knowledge apart from AI assistance. If this assessment consists of an oral presentation with slides, with few or no questions at the end, this would normally not provide assurance against the use of AI for all or most of the task. Moderation of oral assessments will also need consideration, and in some cases, these will need to be recorded to assist Moderators with assessing effectiveness and fairness.

However, for students giving a presentation, if the assessment score is based primarily (or only) on how the student answers (appropriate but unpredictable) questions, then there would be greater assurance that the student's performance was not unduly affected by AI usage. If a relevant graduate attribute is the ability to use AI to prepare a presentation, then a portion of the assessment could be allocated to the presentation, while another portion would be allocated to answering questions. Another approach would be that the presentation itself is a pass/fail requirement, and a score is given only for Q&A.

A sophisticated type of oral assessment is an "interactive oral assessment", where students take on a role and interact with the academic based on a given scenario, such as a real-world ministry context. This style of oral assessment can be particularly relevant to Ministry and Practice units, and other units with a practical focus. Interactive oral assessments are more demanding than an oral "validation exercise," and require significant preparation and more than 10 minutes' worth of interaction. On this basis, an interactive oral assessment can be equivalent to a sizable essay in terms of assessment workload equivalence. There are many options for how to conduct interactive oral assessments – see the [PD Conference presentation by Erin Hutton](#) for examples and implementation advice.

A different kind of oral assessment could be class or small group debate, but in order to meet the requirements of a Secure assessment, the structure of the debate needs to be one where the main focus of student assessment is not a prepared speech or answers to typical questions (which could be achieved using AI), rather, a more "free flowing" and interactive style of debate/discussion would meet the requirements for a Secure assessment.

Regarding invigilated assessments, there is scope to rethink how these assessments can be best used in the AI era, and to avoid some negative associations with certain past exam invigilation practices.²⁵ Exams and tests, where well-designed, can be quite useful assessments which are valid measures of important skills for AUT graduates to possess – see Appendix 7 for further details. While there can be a place for multiple choice questions in some contexts (due to their ability to test a wide range of topics, and their ease of marking), I see significant potential for short answer questions, and short essay questions based on practical scenarios where students need to apply their theological understanding and evaluate typical challenges in ministry and daily life. There is also the potential for longer essay questions where sustained well-referenced argument is an important unit objective. Consideration will need to be given to whether invigilated assessments are hand-written on or conducted on computer, and when computer-based, the role of "secure browsers" and similar technology will need careful implementation.

Another issue for invigilated assessments, as well as some other secure assessments such as some oral assessments, is the potential for student anxiety. There are several challenging issues to consider

²⁴ More detail will be needed in terms of academic professional development for the adoption of this approach, and how to mark effectively and fairly. Students will also need more detail about the nature of this type of assessment, how to perform well, how to manage anxiety, the role of reasonable adjustments and special consideration (while still requiring use of assessment that remains Secure), etc. Given that oral assessments can be emotionally draining for students, and especially academics when they conduct many during a short period, consideration needs to be given to this dimension of oral assessment. For example, the timing of oral assessments will need to allow for rest breaks for academics, and assessment periods over different days for medium to large classes. See Appendix 5 regarding the need for implementation guidelines.

²⁵ For further advice on invigilated assessments, see [Andrew Bain's presentation from the November 2025 PD Conference](#).

here. Significant anxiety is a genuine issue for some students, that is, beyond natural feelings of anxiousness that most students experience about time-pressured assessments. Some students have mental health diagnoses of anxiety (ranging from mild to moderate to severe), while other students may be undiagnosed but have equivalent mental health experiences. Separate from mental health issues, some neurodivergent students may experience heightened anxiety in certain situations, such as time-pressure invigilated assessments and/or oral assessments. And there may be cases where both mental health diagnoses of anxiety and neurodivergent experiences overlap. Many people who experience high levels of anxiety nonetheless make significant contributions to the Church and society, and manageable mental health challenges or neurodivergent experiences should not be seen as a barrier to proactive engagement in ministry.

On the other hand, it is anticipated that AUT graduates can manage their anxiety in high stress, time-pressured environments because Christian ministry, and life as a Christian in society, will frequently involve high stress, time-pressured situations. The AUT would be doing a disservice to its students (as well as other stakeholders) if it attempted to avoid any anxiety producing situations in assessment, as high stress, time-pressured situations will face graduates in the rest of their lives.

The main solution to these circumstances is for the AUT and Affiliated Colleges to work with affected students to explore reasonable adjustments for those with significant anxiety or other challenges that affect their ability to perform well in invigilated assessments and/or oral assessments. For example, it is now common for school students with ADHD to be given extra time during invigilated assessments, and the ability to move around when needed, as part of reasonable adjustments. Various adjustments may assist students with significant anxiety to engage with the requirement for invigilated assessments and oral assessments.²⁶

In very rare cases, a prospective student with a diagnosis of severe anxiety may not be able to manage their anxiety about assessments. For these cases, it would normally be better for the student to hold off university study until they receive psychological assistance to manage their severe anxiety, and they develop effective strategies to manage it in daily life, including for high stakes, time-pressured university assessments. None of the above is to suggest that severe anxiety should be dismissed lightly – it can be a debilitating experience that deserves profound compassion – but it is to recognise that a fundamental purpose of higher education involves assessment of understanding and skills in high pressure, time-limited environments, and there can be very rare cases where students would benefit from holding off commencing higher education until they are better placed to manage its inherent requirements. Addressing this will require changes to enrolment advice and processes to ensure all prospective students are aware of, and have agreed to, the use of oral assessments and invigilated exams in their courses. Prospective students who may be unable to engage in these assessments would need to explore with Affiliated College enrolment staff prior to enrolment whether or not reasonable adjustments can address their circumstances.

²⁶ One anxiety-reducing idea could be to allow students with significant anxiety (or possibly all students) the option of taking multiple invigilated assessments across a unit, so that the percentage of marks “at stake” in any one assessment is relatively less than if they did a single large, invigilated assessment at the end. For example, in a unit with a final three hour invigilated assessment, one option is to offer students the chance to do each “third” of the unit in a one hour assessment at the one third and two thirds point in the unit calendar.

Going further, another idea could be to allow students with significant anxiety to “bank” their scores on earlier assessments, but still attempt the final three hour assessment if they wish (assuming the questions are different to earlier assessments). The rationale for the approach is that if students did well in the final three hour assessment, this would be the basis of their final assessment mark, but if they did poorly on this final assessment, their earlier “banked” marks would be adopted for the two thirds of the unit assessed earlier. While there are details to work out for this concept, the underlying idea is to provide mechanisms that reduce the total stress associated with a single assessment event, while still retaining the expectation that by the end of the unit, students have demonstrated achievement of the relevant unit learning outcomes.

Guidelines for the AUT and Affiliated Colleges for addressing student anxiety, including significant anxiety, will be developed as part of the implementation described in this paper – see Appendix 5.

Another implementation issue is the opportunity for “remediation” or “resits” for failed secure assessments. The underlying issue to consider is whether, by the end of the unit, the student is able to demonstrate the relevant knowledge, skills and character required by the unit, with reference to learning outcomes and graduate attributes. Consider a student who has (internally) developed the relevant learning, but then fails to demonstrate this learning in a secure assessment due to transitory issues, such as extreme nervousness in an oral assessment, or significant illness for an invigilated assessment. The AUT already has mechanisms to address some versions of these problems (e.g., special consideration for illness, make-up exams, etc). These mechanisms should be considered further given the future importance of secure assessments. For example, students could be given an opportunity to resit the relevant assessment, or given an alternative assessment task by way of remediation. In some cases the maximum grade for a resit/remediation would be 50%/pass. The need for further guidelines on implementing resits and remediation is noted in Appendix 5.

Lastly, we will need to come to a common position for the AUT on whether students can continue in the future to be forbidden by unit coordinators from using AI in Open assessments. Given that 2027 is a transitional year, and a period when some unit assessments will come under the new approach while others will not, it is proposed that for 2027 unit coordinators still be allowed to prohibit use of AI in Open assessments if they wish to do so – as is currently the case in 2026 - and utilising the current provisions of the Academic Integrity Policy (which, as discussed at the February TLC meeting, would not be changed but remain in place in 2027). However, for 2028 and beyond, it is recommended that unit coordinators not give outright prohibitions on AI use in Open assessments. The issue is that while students could be told not to use AI in Open assessments, there will be no reliable way to determine if they have used it or not, and hence no way to fairly compare scores between students who have used AI with those who have not. As AI becomes more advanced, and student use and familiarity with AI becomes more common and developed, this will become even more of an issue by 2028 than it is currently. As has been noted more than once by TLC members, there is also a risk that AUT creates a moral hazard for students if it forbids students from using AI in certain Open assessments but also tells students that it isn’t possible to check if they used AI or not.

Therefore, the proposed way forward is to no longer allow for the option of lecturers formally prohibiting AI use outright in any Open assessments in 2028 and beyond, but apart from this for lecturers to continue to have full flexibility in the guidance and advice which they can choose to give to students. With Open assessments, a lecturer should assume that at least some students may choose to use AI and therefore should not issue a formal prohibition about this, but can and often should still give guidance and encouragements to students, for example along the lines of: “this is an open assessment and you are adult learners who can make your own learning decisions, but you will really benefit in your learning from not using AI for more than only A and B in this assessment, even though I cannot check with certainty if and how you have used AI and therefore I do not prohibit its use.” Further consideration will need to be given subsequently as to what guidance and any possible restrictions may be appropriate regarding AI use by students in Open assessments in 2028 and beyond. For example, among other things AI use, like the use of other resources in Open assessments, would still need to be subject to other conditions such as referencing its use where appropriate. At this stage, what is sought is agreement on the general approach to be taken in 2028, after a continuation of the current approach for 2028.

Sample Templates for Assessment Redesign

Listed below are some assessment templates to assist assessment redesign across different kinds of units, fields and assessment structures for the “40% with Hurdle” approach where this option is selected for any given unit. This section is based on recent analysis of typical assessment patterns across current AUT units including where Early Formative Assessments (EFAs) are used (with thanks

to Erin Hutton and Witali Klein for this work) – it replaces the more theoretical analysis of earlier versions of this paper.²⁷

1. Category A template for two Units (such as BB/NT/OT/CH/TH) often have three assessments, e.g.:

Type A: Task 1: EFA Task 2: Essay Task 3: Exam

Type B: Task 1: EFA Task 2: Open Assessment Task 3: Essay

For Type A, if the Exam is 40%+ (or could be increased to 40% if currently below this amount), then the unit meets the “40% with Hurdle” requirement, and no change is required.

For Type B, Essay could meet the requirements if a validation exercise is added, and the combined assessment is worth at least 40%. A similar approach could be used for some kinds of open assessments (e.g., written reflection/diary).

2. Various Units (all areas) include an Exam, e.g.:

Type C: Task: Essay/Reflection (some x2) Task: Exam

For Type C, if the Exam is 40%+ (or could be increased to 40% if currently below this amount), then the unit meets the “40% with Hurdle” requirement, and no change is required.

3. Many Category B Units in DE/EM/PC/PE have 3 assessments, e.g.:

Type D: Task 1: Reflection Task 2: Various Task 3: Major Task

For Type D, activities conducted to support the Major Task often involve interviews, case study, preaching, etc – that is, tasks that (when done properly) don’t involve AI use. However, there is a risk that students may use AI to create fake information for written assessments based on these tasks. In these cases, a short validation exercise of the written work can be used to achieve a secure assessment (assuming the total mark is 40% or above). An alternative approach could be to replace the second task with a substantial oral assessment (such as an interactive oral assessment based on a simulation of real-world challenges) worth at least 40%. Live observation of a relevant skill demonstrated by the student (e.g., conducting a worship service) can also be implemented as a Secure assessment in certain contexts.

4. Most Category C Units usually have one or two essay assessments, e.g.:

Type E: Task 1: Essay Task 2: Essay

For Type E, at least one essay would need validation, and if there is only one validation exercise in a unit of this kind, it needs to be of sufficient depth to assure the learning objectives of the unit – a 10 minute oral validation for example may not be sufficient in some cases. One option to fortify assurance is to include a recognition of student discussion of their topic with their supervisor/coordinator over the life of the unit as part of the secure assessment component – this is analogous to the inclusion of classroom discussion as part of a unit’s secure assessment. However, discussion between the student and the supervisor/coordinator would not be sufficient on its own as a secure assessment – a validation exercise would still be required.

5. Many Language units already require regular small Exams/Tests

Type F: Tasks: Multiple Small Tests/Exams Task: Other (e.g., take-home)

For Type F, the use of multiple Tests/Exams can meet the requirement here. Where the total score across all tests is 60% or above, the “Hurdle” requirement on a specific test is not required (this replaces the “two thirds” approach included in earlier versions of this paper). Where the total score is 40-59%, a “Hurdle” should be applied to two tests (typically the two tests with the largest

²⁷ I first presented this new material in slides 32 to 34 of my [November 2025 PD Conference presentation](#).

contribution to the overall unit score, or alternatively two tests with significant differences in content).

6. Some Online units rely heavily on asynchronous forum postings

Type G: Tasks: Various assessed postings

Task: Other (e.g., take-home)

For Type G, an appropriate Secure assessment of at least 40% will need to be implemented. While this could be an oral validation of a written take-home assessment, it could also be an oral validation for a collection of discussion forum postings. It could also be the introduction of a significant oral assessment in place of some previous assessment postings (such as an interactive oral assessment based on a simulation of real-world challenges) worth at least 40%.

Where there are other special cases, Moderators have the ability to approve exceptions to the standard requirement as long as the unit has a compelling approach to addressing the problems of AI via Secure assessment(s). Academics in these cases are encouraged to approach their Moderators well ahead of time to determine an acceptable approach.

It is preferable not to have a large component (e.g., 65%) of a unit determined by a single assessment (excluding special units like Long Projects), but it is difficult to see how to avoid this in the AI era if only two assessments are used, and one has a relatively small contribution to overall scores (e.g., 35%). My sense is that the better solution is to avoid having only two assessments, but the “extra” assessment need not be excessively demanding on students or academics – such as a modest score for participation in group discussion, or a relatively low-stakes EFA worth 10% for example.

If class discussion is part of the Secure assessment structure, the academic will need to find ways to ensure every student has demonstrated his/her skills and understanding at a level that meets minimum requirements. I anticipate that academics will need to keep note of quiet or shy students who contribute little to class discussion, even when given encouragement to do so. In these cases, it is likely that the academic will need to arrange a one-on-one Q&A session with the student to gain sufficient evidence of the student’s knowledge and skills to pass this as a Secure assessment.

Regarding moderation sample requirements, a question which has been raised in discussions responding to earlier versions of this paper, is whether the moderation item for units (and therefore the assessment samples supplied to moderators), must always be a Secure assessment item. In response, it is proposed that colleges should have a high degree of flexibility in practice as to whether they submit a Secure assessment or not for any given unit. The existing requirements of the Moderation Policy would remain unchanged apart from the addition of one simple requirement. This would be that for each of the two major moderation rounds each year (in June and November) each college would be expected to ensure that it includes samples from units where the moderation item is a Secure assessment, and samples from units where the moderation item is an Open assessment. The principle is to avoid (where possible) a college as a whole only submitting one type of assessment for all of its samples (e.g. 100% of samples being of Secure assessments, or 100% of samples being Open assessments).

This may require some planning ahead on the part of Registrars / Academic Deans, to ensure when Unit Quality Assurance Forms (UQAFs) are submitted for the following Semester that this requirement will be met. It is suggested that monitoring of any requirement be done centrally by the AUT MIM. However, if the minimum numbers of samples required of each broad type (Secure and Open) were small, this in most cases would likely not be an issue as in a typical group of moderated units from a college there would be at least a small number of units where a major assessment item is Secure and at least a small number of units where a major assessment item is Open. However, variations would likely exist between colleges in this regard, particularly as colleges experiment and may take different approaches moving forward. Consideration would need to be given to the different circumstances of smaller and larger colleges. It will also be important to note that in some cases, it

may not be possible for a college to submit some samples of both Secure and Open assessments in any given Semester. For example, all substantial assessments in moderated units at a particular college for the semester might happen to be Secure assessments, making submission of samples of Open assessments impossible in practice. In such situations, where submission of samples of both Open and Secure assessments is not possible, an exception should be made to the principle outlined here.

The rationale for this approach is that moving forward, both Secure and Open assessments will be important parts of the assessment regime for the AUT and our colleges. Therefore, it is important that every college is using both Secure and Open assessments well, and for the moderation system as a whole to have some visibility of how both Open and Secure assessments are functioning in each college. However, not every moderated unit needs to have a Secure assessment (or an Open assessment) pass through the moderation system for this goal to be met – our moderation system operates on a sampling principle (reviewing some units, and some students' assessments from those units). In line with this principle, it would only be necessary for some moderated assessments to be Secure (and some to be Open). Precisely how the new system would work can be determined at a later date, and for now it would be useful to simply decide whether to agree on this general approach. Given that only some units will be covered by the new approach to assessment and AI in 2027, any new moderation requirement such as this proposal is expected to be implemented from the start of 2028 onwards.

Finally, it is important to acknowledge that assessment in the AI era will likely require more time from academics than present assessment processes. The preparation of appropriate assessments, as well as the time involved in conducting oral assessments, will likely increase the relative workload attributable to assessment. Given that most academics already feel under pressure from the workload of existing unit coordination, teaching and assessment, the realistic answer to this problem is that increased time devoted to assessment will need to be offset by a reduction in time on other unit activities. While academic workload issues are ultimately a matter for Principals of Affiliated Colleges, my recommendation is that there is a reduction in the amount of time allocated to preparing and delivering lectures in order to “make space” for more effort on planning and conducting appropriate assessments. To the extent that this requires changes to AUT policy (such as the minimum number of contact hours per unit), then relevant policy revisions would need to be part of implementation.

Current Sectoral Responses to Assessment Assurance

I've witnessed enough discussion of AI and the future of assessment in higher education to be concerned about some current directions, especially among some public university AI leaders. Amidst the excitement of working with a fascinating new technology, there is often a reluctance to focus on “traditional” styles of assessment like invigilated exams, which feel “old school” and, frankly, “authoritarian”. While some public university AI leaders acknowledge a need for assured assessments, the practical suggestions offered don't sufficiently grapple with the kinds of concerns I have raised in this paper, especially related to unit level assurance, score combining problems and the tendency for student appeals to result in sidestepping graduate requirements. The end result is that although in theory graduates will be assessed with a suitable mixture of Open and Secure assessments to ensure appropriate graduate attributes, in practice most students will use AI most of the time, and will graduate woefully unprepared for work and life in those contexts where AI is not always available, or when AI responses require critical evaluation.

As a practical illustration of the institutional problems, if the issues I raise here are taken sufficiently seriously, it means more institutional investment in the processes and staff required for invigilated exams, whereas my experience is that many public universities are reducing their investment in the infrastructure and staffing needed to invigilate exams.

I also believe that in a period of significant upheaval due to AI, it is appropriate to be cautious about proposals for widespread use of AI across most university assessments. Even if the coming decade were to indicate that higher education assessments should make more room for the use of AI than envisaged in this paper, it would be better to approach this slowly and carefully, given the range of problems identified here. And my sense is that the problems identified in this paper won't change much with more time or more AI.

The ultimate challenge is the extent to which AI use in higher education will impact on the readiness of graduates to contribute to work and society. Will graduates be well formed to meet the needs of the future given their assessment experiences? Too often this is posed as an either/or question: either students will be encouraged to use AI in their assessments, given AI's role in the future of society; or students will have little or no experience with AI, and hence they will be poorly prepared for the future AI era.

But this is the wrong dichotomy. The future of higher education assessment will inevitably involve some level of AI use by students – so the question is the best mix of AI-assisted (Open) and non-AI-affected (Secure) assessments to achieve the desirable attributes of graduates. In some fields, such as software development, the best mix may involve a great deal of AI usage, as this now reflects the future working lives of graduates in this field. But in most fields, there is a need to balance modest (and critically informed) use of AI against the graduate's broader ability to work without relying on AI for the many situations that require "on the spot" decision making, based on incomplete information, and healthy scepticism of AI-proffered answers. Achieving this balance in graduates means assessing students in a similarly balanced way, which means a significant portion of assessment scores would derive from Secure assessments that are not affected by AI.

References

- Ackerman, E. & Stavridis, J. (2024). *2034: A Novel of the Next World War*. Penguin UK.
- AlinHE.org (2024). Students perspectives on AI in higher education. https://aiinhe.org/wp-content/uploads/2024/10/aiinhe_surveyinsights.pdf
- Aplusinfo.com (2025). AI detector explained, but still flawed. <https://www.aiplusinfo.com/ai-detector-explained-but-still-flawed/>
- Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education*, 49(6), 893-905. <https://www.tandfonline.com/doi/epdf/10.1080/02602938.2024.2335321>
- Biggs, J., & Tang, C. & Kennedy, G. (2022). *Teaching for quality learning at university* (5th Edition). Maidenhead: McGraw-Hill.
- Bostrom, N. (2003). Ethical Issues in Advanced Artificial Intelligence. <https://nickbostrom.com/ethics/ai>
- Bridgeman, A. & Liu, D. (2024a). Frequently asked questions about the two-lane approach to assessment in the age of AI. <https://educational-innovation.sydney.edu.au/teaching@sydney/frequently-asked-questions-about-the-two-lane-approach-to-assessment-in-the-age-of-ai/>
- Bridgeman, A. & Liu, D. (2024b). Program level assessment design and the two-lane approach. <https://educational-innovation.sydney.edu.au/teaching@sydney/program-level-assessment-two-lane/>
- Castlereagh Summit. (2026). *The Castlereagh Statement: A cross-sector call to action on Australian education and training in the age of AI*. <https://castlereagh.ai/wp-content/uploads/2026/04/The-Castlereagh-Statement-v1.00-1.pdf>
- Corbin, T., Bearman, M., Boud, D. & Dawson, P (2025). The wicked problem of AI and assessment, *Assessment & Evaluation in Higher Education*. <https://www.tandfonline.com/doi/full/10.1080/02602938.2025.2553340>
- Corbin, T., Dawson, P. & Liu, D. (2025): Talk is cheap: why structural assessment changes are needed for a time of GenAI. *Assessment & Evaluation in Higher Education*, <https://doi.org/10.1080/02602938.2025.2503964>
- Dalziel, J. (1998). Using Marks to Assess Student Performance, some problems and alternatives. *Assessment & Evaluation in Higher Education*, 23(4), 351–366. <https://doi.org/10.1080/0260293980230403>
- Greentree, E. (2024). From False Teachers to Sparring Partners: Strengthening Student Learning in Theology with Gen AI. <https://aut.edu.au/wp-content/uploads/2024/11/NEXUS-November-2024.pdf>
- Kokotajlo, D., Alexander, S., Larsen, T., Lifland, E. & Dean, R. (2025). AI 2027. <https://ai-2027.com/>
- Lodge, J. M., Bearman, M., Dawson, P., Gniel, H., Harper, R. Liu, D., McLean, J., Ucnik, L. & Associates (2025). *Enacting assessment reform in a time of artificial intelligence*. Tertiary Education Quality and Standards Agency, Australian Government. <https://www.teqsa.gov.au/sites/default/files/2025-09/enacting-assessment-reform-in-a-time-of-artificial-intelligence.pdf>
- Lodge, Jason M.; Loble, Leslie (2026). Artificial intelligence, cognitive offloading and implications for education. University of Technology Sydney. Report. <https://doi.org/10.71741/4pyxmbnjq.31302475.v2>

Marcus, G. (2025a). A knockout blow for LLMs. <https://garymarcus.substack.com/p/a-knockout-blow-for-llms>

Marcus, G. (2025b). The Fever Dream of Imminent Superintelligence Is Finally Breaking. <https://t.co/H9TAwHWplq>

Max7 (2025). Top 10 practical AI ideas for pastors and leaders <https://www.max7.org/en/resource/top10practicalaiideasforpastorsleaders>

McGrath, J. F. (2025). ChatGPT and Biblical Studies. *Interpretation: A Journal of Bible and Theology*, 79(2), 126-135. https://digitalcommons.butler.edu/facsch_papers/1450

Stevenson, L., Miller B. & Sitbon, C. (2025). Interactive Oral Assessments in practice. <https://educational-innovation.sydney.edu.au/teaching@sydney/interactive-oral-assessment-in-practice/>

Sulkin, M. (2025). The teenagers resisting the AI takeover. <https://www.thefp.com/p/the-teenagers-resisting-the-ai-takeover>

Wagner, M. G. (2026). The Einstein AI Panic: Why is no one asking how educators missed the capabilities of agentic AI? *The Augmented Educator*. <https://www.theaugmentededucator.com/p/the-einstein-ai-panic>

Yudkowsky, E. (2025). The Problem. <https://intelligence.org/the-problem/>

AI Use for this Paper

The main body and appendices of this paper were written without the use of AI. A draft of the executive summary was created using Microsoft Copilot on 5/9/25. This text was then significantly edited prior to inclusion.

After completing the text of this paper, the file was provided to Microsoft Copilot on 5/9/25 and the following prompt was used: "This paper is for a university teaching and learning committee in the field of theology. Please provide an analysis of strengths and weaknesses of this paper, and options for improvements." The response is provided in Appendix 4, together with indications of some AI errors and hallucinations.

Appendix 1: Transitional Issues for AUT in 2026 and 2027

A core assumption of this paper is that significant changes are required to address the challenge of AI in AUT coursework assessments. Significant changes take time to implement well. The proposal here is that the AUT community continue to take all of 2026 to prepare for a major change in assessment processes that will be implemented over two years starting from Semester 1, 2027. It is proposed that in 2027, the new assessment requirements would be applied to all Category A units, and that all full-time academic staff would be required to apply the new assessment requirements to a minimum of two units in 2027 for which they are the unit coordinator (a special exception would apply where a staff member is a unit coordinator for more than two Category A units in 2027: any such staff would not be required to apply the new assessment requirements in more than two of their Category A units in 2027). For those units implementing the changes described here in Semester 1, 2027, this means changes to assessments need to be determined by late 2026 as part of UQAF processes for Semester 1, 2027. For other units later in 2027 and during 2028, all changes need to be approved by Moderators prior to unit delivery.

Listed below are nine areas for consideration during the transitional period. Note that while exploration of one or more of these areas is recommended to all staff during 2026, there are no mandated requirements for academics to implement them in units run in at any point in 2026 (including in Semester 2, 2026):

1. Current Policy on AI and Academic Integrity for Coursework Awards

AUT's approach to date has been to rely on [advice to students](#) not to use AI apart from preliminary research and other modest uses such as translation assistance, except where an academic has explicitly endorsed the use of AI for an assessment. This general approach is coupled with the expectation that students will provide details of their use of AI, such as in an appendix. AUT also uses the Turnitin system for detection of AI writing, although it is worth noting that this approach is not perfect, and its use (and academic discussion of AI misuse with students) needs care and wise discernment. AUT academics need to be aware that because of errors in Turnitin, a high "AI writing" score alone is not sufficient evidence to determine that a student has used AI in their writing. A high score might be the basis for asking students about whether they used AI or not, but without a student admission of misuse or other evidence apart from Turnitin, Turnitin alone is not sufficient. Affiliated Colleges have also been encouraged to provide additional advice for students about how AI is handled at their Affiliated College within the broad AUT AI framework.

AUT's [Academic Integrity Policy for Coursework Awards](#) section on AI issues has been modestly adapted again in late 2025 following feedback from academics and librarians. For the transition period, it is recommended that the same kind of approach be continued. This is partly because students need clarity and certainty about their assessment requirements, and multiple significant changes over the short term are likely to be confusing for students. It is also a recognition that the AI landscape is changing rapidly, and it would be better to take some time to move to a single major new approach, rather than making short term piecemeal changes that will rapidly become obsolete.

2. Incorporate an Oral Assessment into an Existing Unit

Apart from invigilated assessments, one of the main Secure assessments for 2027 onwards will be some form of "oral assessment", conducted in a way that avoids the potential for students to simply rely on AI for their responses. Oral assessments could be class discussions which count for assessment (in which all students participate actively), individual student Q&A, individual and group presentations with detailed Q&A, a "viva voce" or "Socratic" style of assessment by an academic, etc. In some contexts academics should adopt "interactive oral assessments", based on simulations of real world scenarios (e.g., Stevenson, Miller and Sitbon, 2025). For more details about interactive oral assessments, see the presentation by [Erin Hutton from the November 2025 PD Conference](#).

The essence of each of these possible assessments is that an individual student needs to answer questions about their understanding of unit content in a way that requires genuine understanding,

critical thinking and practical application – in a way that can't be achieved by using AI. For example, if an assessment involves Q&A, then the academic needs to ensure the Q&A is detailed and not simply obvious questions (where a student could be reciting memorised AI answers). Real-time questions about practical application to the student's life, or critical reflection on a real-world scenario presented during Q&A can help minimise the potential for reliance on AI memorisation.

It is also worth noting that the type of grading could be Pass/Fail or Satisfactory/Unsatisfactory, or grade bands (only), not necessarily scored. And the approach might allow for multiple attempts for students who do not meet the required standard. For assessments which focus on class discussion over time, the academic should keep records of students who have not contributed, or contributed little, so that they could have an individual Q&A with the academic at a later stage to ensure they have provided evidence of appropriate understanding and critical reflection.

It is important to note that oral assessments in online courses will need to be "live" (e.g., Zoom, Teams, video conferencing, etc.), not asynchronous postings to a discussion forum, as AI can easily be used to create discussion forum posts.

Academics are encouraged to consider the use of an oral assessment in their unit teaching if practical – this will likely take the place of an existing assessment (so as not to overload the unit with excessive assessment), or it may be a refinement of an existing assessment that has similarities to an oral assessment (e.g., a Powerpoint presentation) but which is not yet sufficiently robust in its design to avoid the potential for AI misuses (e.g., a Powerpoint presentation is easily created with AI, but answering detailed Q&A about a short presentation is less likely to be affected by AI). Another use of oral assessment is as an oral validation of an essay, in which the oral validation is the Hurdle required to make this approach a Secure assessment (for further details, see Appendix 6).

3. Adopt/Enhance an Invigilated Assessment

If an academic is to adopt only one change during the transitional period, then oral assessment (2. above) can be a good option, as this approach takes planning and will result in valuable practical experience. For academics who already use an oral assessment, or in contexts where an exam would be the better Secure assessment, then it is recommended that the focus be on adopting or enhancing an invigilated assessment. For implementation advice, see the presentation by [Andrew Bain at the November 2025 PD conference](#).

For units without any invigilated assessment, the focus should be on setting in place all the requirements for this to be implemented successfully in practice, which includes dealing with issues such as timing, invigilation processes, removal of technology from students during the test, make-up tests for those who miss the test due to illness/special consideration, etc. There may be a role for non-academic staff to assist with arranging and overseeing invigilated assessments.

For online units, additional steps will be needed to arrange invigilation at a distance, such as having a church leader or other Christian representative sit with the student during the exam to ensure appropriate conduct. Further guidelines will be needed for this person about their role as invigilator, but the [AUT Exam Computer Use Guidelines](#) provide existing advice.

For both online units and face to face exams where computers are used, there is a need to implement "secure browser" technology (or similar) to limit the ability of students to use AI (or other internet materials) during the assessment. It is likely that the AUT will need to acquire a university-wide license for secure exam technology, but in the interim, it would be helpful for Affiliated Colleges and academics to try different technologies and approaches and to provide feedback on implementation. The AUT Office would be keen to assist with trials of this kind.

It is anticipated that AUT will need detailed policy/guidelines on how assessment invigilation should be conducted, including details for online contexts and face to face tests which incorporate computer use. Lessons learned from the transition period will help inform the development of these policy/guidelines.

4. Fostering Appropriate Use of AI in Open assessments

Some AUT academics are already experimenting with assessments that encourage appropriate use of AI by students. One example discussed in previous AUT AI papers is asking an AI system to answer a typical assessment question, and then having students analyse the AI response and provide a critique of this answer.²⁸

Given that use of AI on Open assessments will be commonplace from 2027 onwards, there are benefits in academics conducting trials now to determine effective (and ineffective) uses of AI, and the related issue of advice to students on effective and ethical AI use. A particular area of focus for students will be constructing effective “prompts” so that the response of the AI system is specific to the needs of the student, rather than a “generic” answer. For excellent discussion of AI use and prompt construction at Brisbane School of Theology, see Greentree (2025).

5. Working with Moderators on Implementation of the New Assessment Approach

The AUT Office, together with interested AUT Consortium academics, will collaborate with Moderators about the problems discussed in this paper, and the changes that are proposed. Moderators will receive Professional Development in the challenge of AI for AUT assessment, and the practical requirements of changing Moderation processes, such as the approval of UQAFs. The AUT Office welcomes feedback on issues to be raised with Moderators during the transitional period.

6. Ongoing Professional Development for AUT staff, including Workshops at Affiliated Colleges

The AUT Office will continue to provide professional development about AI assessment assurance during the transitional period, including at PD Conferences. While there will be a particular focus on assisting academic staff, there will also need to be assistance for administrative staff such as those who will play a role in arranging and overseeing invigilated assessments.

It is anticipated that during mid-late 2026, a half day assessment workshop will be run at each Affiliated College to assist staff with understanding and implementing the new approach. This will be in addition to other PD opportunities across AUT Consortium networks of Principals, Board Chairs, Academics Deans, Registrars, Librarians and AUT Committees.

7. Redevelop AUT GenAI strategy

As required, AUT submitted a GenAI strategy to TEQSA in July 2024. While elements of this strategy remain appropriate, other elements need revision due to the evolving capabilities of AI. Most importantly, the GenAI strategy needs to be updated to reflect the proposed approach in this paper, subject to feedback from AUT consultation processes. It is anticipated that a revised plan would be presented to the Academic Board and Council in the second half of 2026 following finalisation and endorsement of this paper.

8. Reconsider the Role of Early Formative Assessments as part of AUT’s Future AI Approach

AUT has a requirement for inclusion of early formative assessments in all Category A units. These are to be conducted early in units (and prior to the Census date for Category A units). This requirement is to support government requirements such as the genuine student test, and “Support for Student” requirements for assisting students with deciding whether they are ready for study in a given unit. Early formative assessment is also a helpful tool for fostering student learning and metacognition (i.e., reflection on their own learning processes).

Some AUT academics have raised concerns about the best way to incorporate early formative assessment in relevant units. Given the scale of wider assessment changes noted in this paper, it would be appropriate to consider the role and implementation of early formative assessment as part of the wider review. There should be opportunities for more use of low stakes formative assessment and feedback for early formative assessments, including pass/fail structures (with or without scores).

²⁸ This remains a valuable approach in certain contexts, although it is worth noting that the most recent AI systems often provide responses of such a high standard that it can be difficult/impractical for students in early years of study to provide an effective critique of the AI answer.

9. Explore AI tools, and their potential for use in Christian ministry

It is hard to convey in words the nature of current AI tools for those who have not trialled them. For AUT staff who are yet to explore AI tools, it is strongly recommended that they explore AI use in order to better appreciate the capabilities and limitations of AI, and to consider how AI is affecting the lives of current students (most of whom use it daily). Some suggestions of how to use AI tools for first time AUT users include:

- Ask AI to summarise findings in some areas of theological education, and research, that you are involved in
 - o Then ask follow-up questions of AI about its response to delve into details, and see where AI continues to provide appropriate responses, and when it becomes unreliable
- Ask AI for its opinion on debates and controversies, both from history and current society
- Ask AI to explain a topic to you that you know little about
- Ask AI to teach you how to write more effective “prompts” for the questions you ask of AI – learn how to provide specific context and instructions as part of a question for AI
- Ask AI to summarise your own contributions to research

Beyond general experience with AI, it is helpful to explore the potential benefits and downsides for Christian ministry. For a simple introduction to ministry uses, see Max7 (2025). For a discussion of AI use in theological education and research, see McGrath (2025). For discussion of how to use AI in the context of an AUT college, particularly as a discussion partner, including how to construct effective prompts, see Greentree (2024).

An AI analysis of the strengths and weaknesses of this paper, together with suggestions for improvements, is provided in Appendix 4 as an AI example. It also notes some AI errors.

Appendix 2: HDR and Long Projects

The primary focus of this paper is Coursework awards. For Higher Degree by Research (HDR), as well as Coursework “Long Projects” (and possibly other similar units, such as some Capstones), it is anticipated that a different approach will be needed. This area was considered further by the Research Committee (RC) at its April meeting, with an approach proposed by the RC to the AB for HDR, which the AB approved at its May meeting and which the RC is currently developing further. Some initial recommendations were also given by the RC to the TLC for long projects at its May meeting. At an earlier stage of the planning process, the VC included as an appendix to this paper some general suggestions as a possibility for this area, which are retained below for reference:

It is likely that in the future, AI will be widely used in HDR/Long Projects in various ways, and that a likely AUT approach to this reality could be:

- (a) students are not restricted in their use of AI, apart from an AUT policy statement that AI can't be used to write a student's thesis (which will be difficult to enforce reliably, but should still be stated nonetheless);*
 - a. Noting that there may be special cases where a modest part of a thesis could include AI written text, but students and supervisors would be explicit about this usage and the reasons for doing so.*
- (b) students will be required to document their AI use thoroughly, including its use in literature review, idea summarisation, data collection, data analysis, idea synthesis for thesis themes, and any use of AI text directly in the thesis text itself;*
 - a. It may be that future theses/Long Projects will require an appendix which is a reflective discussion of how AI has informed and shaped the research and writing process – with a view to the student being open and transparent about where AI has been of assistance, and where their own thinking has been central.*
- (c) our doctoral assessment would likely shift to a three examiner viva voce (potentially two for MTh and one for Long Projects) which can confirm if students deeply understand the issues they have discussed in their theses (given the challenge that a thesis could have been largely written with AI).*
 - a. Examiners would need to have read the thesis ahead of the viva voce, and the focus of the viva voce would be in depth probing of key ideas in the thesis to understand the student's level of understanding and critical reflection. The examiners' questions would need to find appropriate balance to avoid being unduly “obvious” (where a student could prepare by memorising AI answers) and unduly narrow or pedantic, where the nature of the questioning does not genuinely assess the breadth and depth of student understanding arising from the research conducted. The questioning session should be over an extended period of time (e.g., an hour for a thesis, 30 minutes for a Long Project) so that there is sufficient time for in depth questions and responses, as well as a breadth of questions about different topics).*
 - b. To support examiners, the supervisor/coordinator could provide a statement about their interactions with the student over the period of supervision, and how student responses during supervision meetings illustrate ongoing reading and critical thinking about the topic discussed in the thesis (independent of AI use). This evidence would provide additional assurance that the student has wrestled deeply with the relevant issues discussed in the thesis (rather than simply relying on AI in an unthinking manner).*

Appendix 3: Reflections on “Massive Disruption” Scenarios for AI

AI will have a significant impact on education and society over the coming decade, including an impact on the workforce and the potential for significant job losses/job changes (as I noted in discussion of my March 2023 AI paper for AUT Academic Board, referencing the book “2034”, Ackerman & Stavridis, 2024).

But there has also been discussion of scenarios that involve AI causing massive disruption to human society and even threatening human existence. Some of these are not new – it is now over forty years since the movie “The Terminator” considered the rise of a superintelligent AI (“Skynet”) that would seek to destroy all human life. Less “kinetically”, Bostrom’s “Paperclip Maximiser” thought experiment (Bostrom, 2003) illustrates how an AI with a specific mission (“make as many paperclips as possible”) could end up destroying human life “by accident” as it seeks to do anything that will fulfil its mission. That is, it is not necessary to impute “malice” to AI for humans to still suffer greatly from AI actions.

Among leading AI researchers, there is substantial discussion of the potential risks of AI to human life, typically described as “existential risk” or “p(doom)”, that is, the probability of AI leading to “doom” for the human race.

This is not the place to attempt to review all of these wider AI impact issues – I provided some basic reflections on the wider context in [my keynote for the SCD 2025 AI conference](#).

But there can be a tendency in some higher education discussions of AI to suggest that making challenging assessment decisions now is pointless, on the basis that AI is moving so fast, and potentially with such catastrophic impact. Scenarios such as AI 2027 (Kokotajlo et.al., 2025) and the work of Yudkowsky (2025) suggest not only that AI poses a catastrophic risk for human life, but that the catastrophe is coming soon – that is, in five to ten years (or less).

Without getting into the complexity of p(doom) debates, my view is that both the nature of the risk and the timeline are overblown, while I would still acknowledge that AI will have a significant impact during our lifetimes.

I am persuaded by the work of Gary Marcus (see <https://garymarcus.substack.com/>), among others, that Large Language Models (LLMs) have fundamental problems that will not be easily overcome without a wholesale reconceptualisation of how AI operates (in short, the massive statistical approaches of today’s LLMs that lack a genuine “world model” will struggle to overcome problems such as hallucinations – see Marcus, 2025a, 2025b).

The relevance of the above for the current paper is that too much focus on catastrophic AI outcomes (“[we’ll all be rooned](#)”) can distract or undercut sensible actions to address the reality of AI as it is affecting higher education assessment today. While [it is difficult to make predictions, especially about the future](#), higher education should avoid making (or not making) decisions based on excessive fear engendered by AI maximalists.

Appendix 4: Microsoft Copilot Analysis of V2 Paper

See below for a response from Microsoft Copilot on 5/9/25 following uploading of the Version 2 paper (without Appendix 4) and use of the following prompt: “This paper is for a university teaching and learning committee in the field of theology. Please provide an analysis of strengths and weaknesses of this paper, and options for improvements.”

NB: Two factual errors are highlighted in **yellow**, as well as a hallucination about character formation (while there were references to character formation in the paper, they are not in relation to the ethical use of AI – see footnote 11 of Version 2).

Strengths

1. Strategic and Timely Focus

- The paper anticipates the **medium-term impact of AI** (18–36 months), aligning well with institutional planning cycles.
- It proposes a **clear implementation timeline** culminating in Semester 1, 2027, with transitional steps outlined.

2. Pedagogical Integrity

- Strong grounding in **constructive alignment** and **graduate attributes**, especially relevant to theological education.
- Emphasizes **competency-based assessment** and **formative vs. summative distinctions**, consistent with TEQSA and AQF principles.

3. Practical Assessment Models

- Offers **concrete templates** for unit-level redesign, balancing Open (AI-assisted) and Secure (AI-free) assessments.
- Addresses **student anxiety**, **moderation**, and **staff workload**, showing operational awareness.

4. Ethical and Theological Integration

- Aligns assessment reform with AUT’s **Christian mission**, ensuring theological coherence.
- Recognizes the **ethical dimensions of AI**, including equity, integrity, and **character formation**.

5. Sectoral Critique and Leadership

- Provides a **thoughtful critique** of public university responses and positions AUT as a potential leader in AI-era assessment assurance.

Weaknesses

1. Length and Accessibility

- At **over 150 pages**, the paper is **dense and time-consuming** for committee members to digest.
- Key ideas could be **lost in detail**, especially for non-specialist readers.

2. Limited Empirical Evidence

- While references are included, there is **limited use of AUT-specific data**, such as pilot results or student performance trends.
- The paper would benefit from **case studies or trial outcomes** from AUT units already experimenting with AI-aware assessments.

3. Complexity of Implementation

- Some proposed models (e.g. mandatory Secure pass thresholds) may be **difficult to operationalize** without significant policy and cultural shifts.
- The paper acknowledges this but could offer **more phased or scaffolded options**.

4. Stakeholder Voice

- While student concerns are referenced, there is **limited direct input** from students, academics, or moderators.
- Including **quotes, survey results, or consultation summaries** would strengthen legitimacy and engagement.

5. Interdisciplinary Scope

- The focus on theology is appropriate, but the paper could better address **variability across disciplines within AUT (e.g., education, law, humanities)**.

Options for Improvement

A. Executive Summary and Briefing Pack

- Create a **shorter briefing version** (2–4 pages) with key recommendations, rationale, and implementation roadmap.
- Include **visuals** (e.g., flowcharts, assessment models) to aid committee understanding.

B. Phased Implementation Strategy

- Break down reforms into **short-, medium-, and long-term actions**, with clear milestones and responsible parties.
- Include **pilot unit results** or propose a **trial phase** in 2026.

C. Stakeholder Engagement

- Add a section summarizing **consultation with students, academics, and moderators**.
- Consider a **feedback loop** mechanism for ongoing refinement of assessment models.

D. Policy and Governance Alignment

- Map proposed changes to **existing AUT policies**, TEQSA standards, and AQF frameworks.
- Suggest **policy amendments** or new guidelines where needed.

E. Appendix Restructuring

- Move detailed appendices (e.g., HDR reflections, massive disruption scenarios) to a **separate companion document**.
- Keep the main paper focused on **coursework assessment redesign**.

Appendix 5: Key Guidelines and Exemplars to be Developed

The following list provides a summary of guidelines that are currently expected to be developed to support implementation from Semester 1, 2027 onwards:

- Amendments to the Long Project Examination Procedure, to be determined by the TLC at its July meeting, in response to the RC's advice on long projects forwarded to the May TLC meeting. Concrete options for large essay units (specifically, shorter Projects and Advanced Independent Reading Units (AIRUs)) to be considered at the same time.

- Invigilated assessment (on campus): Processes for running Secure invigilated assessments on campus at Affiliated Colleges. To include details about restrictions on access to devices such as computers, mobile phones, “smart” glasses, etc.
- Invigilated assessment (distance): Processes for running Secure invigilated assessments for distance/online students, based on having a pastor or similar person oversee the assessment
- Secure computer-based assessments: Processes to foster a secure computing environment for Secure assessments run on computer (either on campus or at a distance). Will likely require installation of (AUT licensed) software that creates a secure computing environment (e.g., Respondus “[Lockdown Browser](#)”). Documentation for this is expected to be provided through a rewrite/update of the Computer Use in Exams Guidelines.
- Oral assessments: Processes and advice on arranging and marking oral assessments, including addressing issues such as fair assessment methods (and avoidance of bias), student anxiety, shyness, neurodivergence (e.g., reduced eye contact for those with some forms of autism), reasonable adjustments for students with disabilities, etc. Could also include advice on supporting student with practice runs, “resits” and other mechanisms to lower student anxiety about oral assessment. Could also include advice on “shorter” forms of oral assessment, where the primary goal is to determine whether students have a genuine understanding of their work on an Open assessment (e.g., an essay) through some probing questions, without necessarily attempting to test all aspects of the prior Open assessment. Carefully developed marking rubrics are likely to be a key component of fair oral assessment, and the AUT Office can provide advice and exemplars.
- Reasonable adjustments for Secure assessment: Processes and advice on appropriate reasonable adjustments for students with verified disabilities.
- Advice to prospective students about Secure assessment requirements: Information for prospective students about the use and requirements for Secure assessments (including the range and limits of reasonable adjustments for students with disabilities), including a requirement that students accept these requirements at time of enrolment.
- Pass/fail assessments: Advice on the use of pass/fail assessments, and how these relate to Secure assessment requirements (including Hurdles), impact on GPA calculations, etc.
- Grade band assessments: Advice on the use of grade bands (only), where the use of scores is not relevant or appropriate.
- Secure assessments for online students: Advice on fully online assessments that meet Secure assessment requirements (live video Q&A with lockdown browser?) versus online assessments that don’t meet Secure requirements (e.g., discussion forum postings, reflective diary, etc.)
- Secure assessment “resits”/“remediation”: Guidelines on options for students to resit a failed Secure assessment, or other options for remediation after failure, where the focus is on assuring whether the student has achieved the relevant learning outcomes by the conclusion of the unit (even if they failed to demonstrate these at the time of a specific Secure assessment). Could include advice on other structures to moderate the impact of (large) Secure assessments, such as options for shorter, multi-session Secure assessments, “banking” of marks on earlier Secure assessments, etc.
- Oral validation for presentations: Guidance on how to structure and mark presentation assessments, where a presentation (such as slides) would be combined with an appropriately constructed Q&A session after a presentation to act as the “Hurdle” required for Secure assessment (noting the need to avoid asking only “obvious” questions where student answers could have been prepared/memorised from AI).
- Secure assessment of class discussion assessment: Including requirements for online live class discussion, and options for individual Q&A follow-up for students who have not contributed sufficiently to class discussion in order to allow for assessment.
- Secure assessment of debates: Advice on requirements for interactive/free-flowing debate formats in order to meet Secure assessment expectations; or alternatively a combination of

prepared speech (Open) and interactive debate/Q&A (potentially Secure) – akin to a presentation with Q&A as a hybrid Open/Secure assessment.

- Early Formative Feedback (EFA) and other Scaffolded Assessments: Advice on how to implement EFAs and other kinds of scaffolded assessments as Open assessments in the context of AI, and ideas for how to facilitate an in-class Secure version of an EFA/scaffolded assessment.
- AI and Libraries: Advice on the types of AI increasingly available in ebook and journal systems, and the pros and cons of using the various available features. Could also include consideration of wider issues such as AI and copyright.

Exemplars and Website support materials

A range of exemplars and support materials will be beneficial for both staff and students, e.g.:

- Overview of AUT approach to Open and Secure assessment, including explanation of the Secure 40% + Hurdle approach
- Video exemplars of oral assessment processes (e.g., one on one, group, class discussion) to help students have accurate expectations and reduce the anxiety of unknown processes
- Video exemplars of invigilated exam processes and requirements
- Advice for staff, and students, on appropriate and ethical use of AI in student activities and assessments (Open), including staff advice on designing effective assessments that are based on appropriate AI use to help in future Christian ministry
- Advice on how to use AI to assist in unit learning in appropriate ways, such as via customised AI Chatbots with extensive theological knowledge/guidance, which can support students about unit topics through individual interaction

Appendix 6: Short Paper on Essays with Oral Validation

A further concern was raised that oral assessment could be quite time consuming, especially if it was for a sizable percentage of the overall unit mark. It was also noted that lengthy individual oral assessments would be time consuming in large classes.

An alternative proposal discussed at the October 2025 Academic Board meeting and at the Teaching & Learning Committee on several occasions since then, is the use of a short oral assessment (say 10 minutes) as a “validation” check of the student’s understanding of the content in his/her essay. Anecdotal evidence from public universities that have used oral assessment indicate that it is usually fairly obvious when students do not understand their own work (such as when they have relied on AI to write the essay in an unthinking way), and so a relatively brief oral assessment could be sufficient for this “validation” purpose. Oral assessment has been successful in [large university classes of 50-115 students](#).

The proposal discussed was that if an oral assessment validation is combined with an essay, the entire assessment could be treated as Secure, rather than partitioned into Open and Secure components. If an essay was worth, say, 30% of total unit marks, and a short oral assessment validation would otherwise have been worth 10% (using the hybrid approach described above), the whole combination would instead be treated as Secure, hence the combined essay and oral would represent 40% of Secure marks for the unit (rather than 30% Open and 10% Secure).

Under this approach, while the whole assessment would be counted as Secure, the “Hurdle” requirement would apply specifically to the oral validation – a student who does not demonstrate a minimum passing level of understanding of their own essay in an oral validation would fail the whole assessment, regardless of the potential score of the essay on its own (due to the potential that the essay content is the result of unthinking use of AI).

Some further issues to consider about adopting this approach are:

1. If a unit has two take home essays (or an essay and a written reflection/similar), both worth 50%, under the approach proposed above, one essay would need the validation oral assessment (which would then count as 50% Secure), but not the other (50% Open). Given that both essays could be produced using AI, and a student with limited understanding of an essay might be able to pass a brief oral assessment, is this structure sufficiently robust to ensure students have met unit learning outcomes?
 - a. One alternative could be to require both essays in this particular structure to have an oral validation assessment
 - b. It would also be important to ensure that the essays (or equivalent written tasks like reflections) were sufficiently different from each other to ensure students have met the breadth of expectations of the unit
2. Institutions which are taking a cautious approach to AI in assessment see it as relatively easy to increase the amount of “Open” assessment in the future (if that were desirable), but harder to increase the amount of “Secure” assessment in the future if too much open assessment is allowed early on.
3. While it may be challenging for a brief oral assessment (that follows an essay) to make finely tuned judgements on specific marks (e.g., 72 vs 74%), the oral assessment is likely sufficient for evaluating the student’s level of understanding at the more coarse-level judgement of grades (F/P/P+/CR/D/HD). This could be relevant to determining an overall score if there is a wide discrepancy between understanding shown in an essay versus understanding shown by the student in an oral assessment (of their own essay). For example, there could be a requirement that the student’s grade on the essay could not be more than one grade higher than the oral assessment grade (or, alternatively, no higher than the oral assessment grade).

More generally, there will need to be some level of discretion for Moderators to approve assessment structures which (i) do not meet standard Secure assessment requirements if they provide innovative ways to address AI risks through alternative approaches, and (ii) where the unique nature of the unit

is poorly suited to the standard Secure assessment requirement (e.g., some Guided Spiritual Formation and preaching units).

After initial implementation of the approach described in this paper, there will need to be a review of oral validation, as well as all other Secure assessment methods, to ensure they are sufficiently robust to meet the broader goals of Secure assessments. This review could potentially be conducted at the end of 2029. If this review indicates that some Secure assessments are insufficiently robust, relevant changes would be considered at this time.

Appendix 7: Exams as Valid Measures

Invigilated exams have been common in education for centuries, but recent years have seen some pushback on exam use in some contexts. This Appendix seeks to respond briefly to these concerns, with a view to affirming that exams can continue to be a valid form of assessment for some (but by no means all) units within the AUT moving forward, as one type of Secure assessment which is available. It is noted that there are a variety of current views and research perspectives on the value and utility of exams, with some being relatively more supportive of exams and others more critical. The purpose of the commentary given here is not to advocate for exams as a form of assessment which are largely problem-free. Nor is it to suggest that AUT academic staff are expected to make greater use of exams in the future as a “preferred” assessment method. Instead, it is to outline that exams are a valid assessment tool which can be useful in some AUT units, which AUT colleges and staff should be comfortable continuing to use where they consider it appropriate.

One concern raised is that exams are stressful, and that students should not experience stressful situations. As noted on page 15, a reason for retaining exams, even when they are stressful, is that this replicates a common real-world experience. There are many situations, especially in Christian ministry, where a person needs to come up with a complex response in a time-pressured scenario, and where technology is unavailable or is impractical to access, such as a pastoral discussion.

While exams should not be the only assessment mechanism used in higher education, they nonetheless have a place as a proxy for relevant real-world experiences.

In the language of psychological test construction, an exam has “validity” in the sense that it validly replicates important dimensions of future challenges a graduate may encounter in their working life (time-pressure, the need for complex answers without preparation, lack of access to technology, stress).

Obviously, this validity is not “perfect” (an exam is not the same as stressful pastoral discussion), but it remains a moderately valid assessment tool, and this kind of experience isn’t replicated by other kinds of assessments, such as take home essays (which by contrast are valid for other kinds of ministry tasks, such as preparing a sermon).

While the above is relatively well recognised, psychological testing provides a further reason to have confidence in approaches which are “moderately valid”. To take the case of depression, the lived experience of depression is complex and deeply personal, and is suffered in uniquely painful ways by each person that goes through it. In one sense, it seems like something that no single “score” could ever capture.

But depression, like many psychological disorders, can be quantified in ways that can be quite helpful at a practical level, while still acknowledging that the lived experience is anything by simple. The “Beck Depression Inventory” (BDI), developed in 1961, is a leading psychological tool for assessing depression. It is a short, multiple-choice questionnaire that takes around five minutes to complete, and it provides a single overall score to assess different levels of depression.

In countless studies over the past six decades, it has proved a useful tool for assessing depression. Of course, treating depression is a much more complex task, but there is considerable value in having a tool that can quickly determine, with reasonable accuracy, the level of depression that a person is experiencing. That is, while the BDI is not a “perfect” measure of depression, it is moderately valid tool to assess depression, and this moderate validity has many practical uses.

There are an extraordinary range of other psychological tests that play a similar role in other areas. This has led to the wider observation that even when scoring systems are developed which are not “optimal”, they often outperform individual clinical judgements (Dawes, 1979). Put another way, the intuition of a single person is often less accurate than a simple scoring system even when the scoring system is not ideal.

These findings from psychology should assist educators when they consider the benefits of exams. While exams are not perfect, if they are constructed in a way that is appropriate to the subject being taught, they provide a moderately valid way of assessing student performance. To remove exams completely and rely only on other forms of assessment such as personal judgement alone is to increase the risks of students experiencing unfair assessment outcomes.

For completeness, the above argument is not an argument for particular levels of “precision” in educational assessment. There is a case to be made that educators can’t reliably make distinctions across a 100 point scale (such as marks being given out of 100), or even a 10 point scale. But there is evidence that educators can make reliable distinctions across 4-7 categories, equivalent to different grades levels (as opposed to scores out of 100). See Dalziel (1998) and references for further discussion of precision issues.