External evaluation of the Queensland Certificate of Education (QCE) System

Faculty of Education

THE UNIVERSITY OF MELBOURNE

The Assessment & Evaluation Research Centre (AERC)

AUTHORS:

THERESE N. HOPFENBECK, CATHERINE MCCLELLAN, BARRY MCGAW, ZHONGHUA ZHANG, KEIKO BOSTWICK, VINCENT LIANG, & JOSHUA MCGRANE

Acknowledgements

The Assessment & Evaluation Research Centre (AERC) would like to express sincere appreciation to the many individuals and organisations who generously contributed to this evaluation of the Queensland Certificate of Education (QCE) system.

Firstly, thank you to all the Queensland Curriculum and Assessment Authority (QCAA) staff for their collaboration and informative meetings throughout the process, which have ensured that this external evaluation has been held to the highest standards.

We are grateful to the Queensland teachers, school leaders and principals who provided invaluable insights through detailed submissions, comprehensive survey responses, and thoughtful focus group discussions. Their frontline perspectives have been crucial to understanding the nuanced aspects of the QCE system.

Our thanks also extend to the diverse stakeholder groups who enriched our research, including Teachers' Unions, Principals' Associations, Parents' Associations, University Sector (Pathways), Industry Groups, VET/VETDSS Providers, and School Sector Representatives. Each group brought unique perspectives that deepened our understanding of the QCE system.

We are appreciative of the external Advisory Group who provided scholarly guidance throughout this evaluation. This includes Professor Barry McGaw (Honorary Professorial Fellow), Professor Janet Clinton, and Emeritus Professor Field Rickards (Honorary Professorial Fellow). Their independent expertise and critical insights were instrumental in ensuring its rigour and comprehensiveness.

Finally, we extend our appreciation to our research team that were responsible for data collection and analysis in the different phases of this comprehensive evaluation: Pam Robertson, Debbie Krapywnyj, Ruth Aston, Shelley Gillis, Rebekah Luo, Sundance Zhihong Sun, Hayley Paproth, and Oskar Weimar.

Executive Summary

Evaluation Purpose

The University of Melbourne's Assessment & Evaluation Research Centre (AERC) was commissioned by the Queensland Curriculum and Assessment Authority (QCAA) to conduct an independent evaluation of the new Queensland Certificate of Education (QCE) system and its processes. The QCE is the senior secondary schooling qualification in Queensland. Its purpose is to recognise a significant amount of learning achieved by students in the senior years. This recognition is designed to be flexible, broad, and aligned with different post-school pathways, including university, vocational education and training, and work. The QCE is internationally recognised as evidence of student educational achievement by employers, training providers, and institutes of higher education. Learning that contributes towards the QCE can come from a wide range of sources. These include school-based subjects (General and Applied), vocational education and training (VET), recognised studies, and university subjects undertaken while at school. Through this, the QCE represents a move away from a narrow 'one size fits all' senior certificate to a broader, more inclusive model of senior schooling that recognises different forms of learning and prepares students for multiple pathways into work, training, or university.

Legislation in 2006 paved the way for a new senior secondary qualification, and by 2008 the first QCEs were issued. After an independent review by Matters and Masters (2014), from 2019 onward, a revised QCE system was introduced that integrated key recommendations from the review, including changes in senior assessment and tertiary entrance (e.g., new external assessments, stronger quality assurance) under the oversight of the QCAA. The purpose of the new system is to provide students with flexible learning pathways through contemporary, high-quality syllabuses and an integrated assessment program. This program balances internal and external assessments against clear standards, while allowing local context and culture to be incorporated and ensuring comparable results across the state. The system relies on partnership between QCAA and teachers to act to high standards of professionalism, including ongoing professional development, to deliver student credentials that are credible, reliable, and valuable.

The current evaluation considers whether the development and implementation of the new QCE system is effective in delivering equitable outcomes for students. Moreover, it evaluates whether the current processes are sustainable, including making recommendations for future improvements.

Methods

The evaluation encompassed a comprehensive approach, drawing on multiple sources of evidence. These included stakeholder consultations, in-depth analyses of a sample of student outcomes and performance data provided by QCAA, direct observations of key assessment processes and activities, targeted surveys and interviews with diverse stakeholders such as students, teachers, parents, school leaders, QCAA staff and associates. To enhance the rigour of the findings, the evaluation incorporated a targeted review of relevant literature and benchmarking against established best practice criteria, thereby ensuring a robust and triangulated evaluation.

Findings

The evaluation findings suggest that with the implementation of the new QCE, QCAA has established a quality curriculum, assessment, and certification system. The curriculum and syllabuses are clear and well-constructed with extensive consultation and there is already a readiness to review and refine in the light of experience in implementation. The incorporation of learning beyond traditional classroom contexts was found to be valuable. However, there is a need to review the credit awarded for VET Certificate and Diploma courses, particularly in comparison with subjects developed by QCAA.

Stakeholder perceptions of the new QCE system are positive overall. This has supported the successful implementation of the system across the diverse state of Queensland. Numerous channels provide access to stakeholder groups including the QCAA website, professional learning events on a range of topics, communications to schools about specific issues and activities, a redesigned secure portal and newly developed custom software applications to support broad access to extensive, detailed information.

Psychometric analyses demonstrated that QCE assessments produce consistent results that support valid inferences about student achievement. Score reliability consistently exceeded acceptable standards for high-stakes assessment, with minimal bias observed across demographic groups. Quality assurance processes effectively maintain marking comparability, and the combination of internal and external assessments provides complementary evidence of student achievement.

The complete summary of findings from this evaluation is presented below.

System Strengths

The QCE system demonstrates several strengths in design, delivery, and quality assurance. The design balances personalised, contextualised learning with reliability and comparability of results through QCAA syllabuses that clearly delineate the required content while leaving the specific approach to instruction under local control. The syllabuses are standards-based and specify detailed achievement benchmarks, objectives, and assessment criteria aligned with senior secondary schooling outcomes. This level of detail promotes transparency and consistency in instruction, teacher judgment, and student feedback.

Syllabuses focus on cognitive complexity and higher-order thinking, using a cognitive verb framework (e.g., analyse, evaluate, justify) to promote critical and analytical thinking. They are provided for both General and Applied subjects, with integration of VET, recognised studies, and extension subjects. Teacher professional input is included using Subject Expert Panels and teaching networks to contribute to syllabus development and review. This approach increases practitioner ownership and supports contextual adaptation to local needs. Moreover, QCAA's syllabus development and management policy aligns well with best practices internationally, emphasising transparency, stakeholder consultation, monitoring of enrolments, completion data, feedback, and revision of syllabuses.

The system supports teachers through professional development that enhances teachers' assessment literacy skills, supporting them to develop summative assessments that can appropriately incorporate the students' context and culture. Reliability and comparability are maintained through rigorous quality assurance processes, including:

- review and endorsement of the alignment of all summative assessments for Units 3 and 4 with their syllabus completed by trained QCAA assessors
- review and confirmation of the accuracy of a sample of marked student responses against common marking guides completed by trained QCAA assessors
- the incorporation of external assessments developed by QCAA expert staff and administered under common conditions

These processes have been developed to instantiate the previous independent review's recommendation to strengthen senior secondary assessment practices in Queensland (Matters & Masters, 2014).

Opportunities for Improvement

As with any complex system comprising multiple processes and components, and numerous participants and stakeholders, the evaluation identified opportunities to further improve the QCE system design and delivery. These include:

- Developing a comprehensive research agenda to identify factors that might inform revision or adjustment of the current QCE design and delivery, with a focus on monitoring implementation and its impact on delivery conditions, timelines, and teacher and assessor workload
- Expanding external representation in QCAA governance
- Establishing a measurement advisory board with an emphasis on providing guidance on technical and psychometric issues
- Investigating the extent to which experiences with the Access arrangements and reasonable adjustments (AARA) process vary across different student or school factors
- Reviewing the assessment literacy modules offered to teachers to ensure they are providing the most effective and up-to-date guidance for educators
- Increasing use of psychometric analyses to monitor the new QCE system and maintain public confidence in the fairness and accuracy of subject results
- Augmenting the monitoring and quality assurance of marking
- Enhancing the processes used in grade-boundary setting with statistical evidence and clearer documentation
- Improving targeting of assessments to ensure sufficient coverage of the full ability range of student performances

The complete summary of recommendations from this evaluation, including the six broad strategic areas for improvement of the QCE system derived from them, are presented below.

Conclusion

QCAA has developed a distinctive system for the new QCE, one that effectively balances regulatory requirements with local institutional autonomy, supported by a comprehensive quality assurance framework. The design has been implemented with fidelity to its intended purposes and is reinforced by extensive documentation that assists new participants and maintains consistency among experienced practitioners.

This evaluation has also identified areas where QCAA could consider more substantial changes to QCE design and operational delivery. Addressing these areas will require proactive strategies to manage complex, long-term systemic change and ongoing workforce pressures in the education sector. A rigorous, well-documented system monitoring framework, strengthened through enhanced governance, coordinated research and evaluation, sustained workforce capability, transparent documentation, and robust psychometric and technical advice, would assist the organisation as it moves forward in a rapidly evolving education environment.

Summary of findings

The following summary of findings is presented in the order that they appear throughout the body of this report.

Finding	Report Section
Foundation concepts and system development	
• The QCE and QCIA Handbook successfully provides clear, high-quality information about why the new system was introduced and explains its key policies and procedures. The accessibility and quality of policy procedure information appear to meet the needs appropriately for teachers and school leaders, who are the target audience for these policies	Section 3.1
• The current documentation may not be sufficiently accessible for some parents, students and the broader community who also need to understand the system. More detailed information about how assessment processes are designed and developed could benefit some stakeholders. This would help clarify the underlying goals and educational philosophy that drive the QCE system and its supporting processes	Section 3.1
• The recommendations of the Matters and Masters (2014) review have largely been adopted either in full or part in the Queensland Government policy position and have subsequently been implemented by QCAA.	Section 3.1
• Consultations with stakeholders provided information that the program was successfully delivered throughout the challenging times during the pandemic, but the perceptions of the success of delivery varied across the ten school cases	Section 3.2
• Survey data demonstrated a strong perceived understanding of the new QCE system among stakeholders. This finding was corroborated through consultation meetings with teachers and school leaders across all ten case studies, triangulating the survey results and confirming their reliability	Section 3.2
• The QCAA demonstrates a strategic commitment to continue being a learning organisation that attracts and retains skilled professionals through continuous professional development, strategic workforce planning and a robust, inclusive organisational culture	Section 3.2
Curriculum	
• There is strong general approval of the quality of the new syllabuses and, with applied subjects, of the alignment with industry needs. They have a sound base in research findings	Section 4.3
• The one clear exception to this general approval was 2019 Essential English, which did not align well with the other English subjects. It was constrained by two requirements: building on students' Year 10 achievement levels developed with the Australian Curriculum, and seeking to ensure students reach at least Level 3 of the Australian Core Skills Framework (ACSF; McLean et al., 2012)	Section 4.3
• QCE curricula satisfy the quality criteria recommended by OECD. QCAA's syllabus development involves systematic, wide communication with subject and teaching experts and the broader community, reflecting best practice in incorporating advances in knowledge	Section 4.3
• The wide subject offerings of the new QCE system allow students the possibility of pursuing their chosen pathways with the SET planning process effectively, supporting discussions with teachers and decisions by students and their families	Section 4.4
• The duplication policy preventing students from earning multiple credits for the same learning outcomes in different courses is well-regarded	Section 4.4
Assessment processes	
• Endorsement of the IAs assures assessment of the construct as defined in the syllabus and that student responses can be evaluated against the same marking guide standards	Section 5.2
• All assessments are endorsed by QCAA prior to being administered to students in classrooms	Section 5.2
• Endorsement rates have been increasing steadily over time although rates at first submission remain relatively low	Section 5.2
• Annual subject reports contain data regarding the validity and accessibility priorities that led to non-Endorsement of assessments	Section 5.2

Finding	Report Section
• Each syllabus provides detailed information and guidance on the assessment objectives, specifications, conditions, mark allocation, and ISMG for each IA that supports the development of instruments that will fit the attributes of quality assessment	Section 5.2
• Endorsed IAs can be expected to contribute solid evidence to support the validity of inferences about student achievement	Section 5.2
• Confirmation assures the consistent use of the ISMGs across endorsed assessments to assign marks to student responses	Section 5.3
• Cohort Confirmation rates have been slowly increasing over time	Section 5.3
• Provisional mark agreement rates are very high, above 90%, and have been increasing over time	Section 5.3
• The QCE EA and EAMG development process are high quality, well controlled and parallel some of the most highly regarded assessment programs in the world	Section 5.4
• There are several well-defined roles in the development process with clear and consistent documentation of the responsibilities of each role	Section 5.4
• The EAs generally have high reliability and limited error of measurement	Section 5.4
• The correlation between the IA and EA results are moderate to high across subjects, indicating consistency in the underlying construct measured	Section 5.4
• QCAA has implemented annual quality assurance procedures for a sample of Applied and Applied (Essential) subjects	Section 5.5
• QCAA provides advice support and intervention to schools whose assessments are not fully aligned with the specified policies and procedures	Section 5.5
• QCAA does not provide direct follow-up on advice given as part of applied quality assurance, and schools are not required to provide evidence of assessment modification based on such advice	Section 5.5

• QCAA provides extensive assessment literacy training through 3 core and 4 specialised modules, plus additional professional development materials in the Learning Hub. These modules have a range of strengths:
  • Assessment literacy training is easily accessible to all Queensland educators through the QCAA portal
  • The training provided for specialised assessors is engaging, practical, valuable and well aligned to the operational requirements of assessor roles
	Section 5.6

• There are opportunities to improve the assessment literacy training:
  • Assessment literacy training is not mandatory for Queensland educators although the core modules are recommended for all educators
  • The content of the assessment literacy training modules is not well aligned with current thinking in the field of assessment
  • The limited increase in endorsement and confirmation rates over time does not provide convincing evidence to support inferences about the impact of the current assessment literacy training for educators on their enacted assessment development practices
  • The content of the assessment literacy training modules is neither adequately aligned to educators' individual assessment practices nor sufficiently practical to have a meaningful impact on educator's assessment practices.
	Section 5.6

• Incorporation of a module on universal design of assessment systems into the core assessment literacy modules would enhance their value:
  • The course including content on universal design for learning (UDL) in the QCAA Learning Hub is not part of the core assessment literacy training
  • The UDL course content is focused on universal design for learning but is titled "Providing accessible assessment tasks"
	Section 5.6

Finding	Report Section
• The QCE system delivers a unique approach to quality assurance of secondary education with a functional and sound system that delivers quality results	Section 5.7
• Across all processes there are high levels of evidence of each of the attributes of quality assessment provided from one or more of the procedures, supporting its fitness for the current purpose	Section 5.7
• The weighting of IA and EA differs across content areas in line with Queensland policy	Section 5.7
• QCAA provides accessible training to support educators interested in applying for casual assessor roles in the QCAA portal	Section 5.8
• The specialised training for assessors who have completed the online modules and passed the assessment is conducted by expert staff, and is thorough and engaging	Section 5.8
• Assessor training effectively develops applied skills through numerous embedded exercises	Section 5.8
• Those in casual assessor positions feel that they are well prepared for their roles and that they had adequate time and resources to complete their duties	Section 5.8
• Stakeholders in the QCE system expressed concerns about the long-term viability of the casual assessor workforce	Section 5.8
• QCAA provides extensive documentation of all operational procedures and roles in the QCE processes	Section 5.9
• The structure and content of the QCAA documentation are presented in a consistent and accessible format	Section 5.9
• AARA information available is comprehensive and easily accessible online, and support is available if needed	Section 5.9
• End users of the AARA process and procedures have mixed experiences	Section 5.9
• No statistically significant bias in results was found between students who receive assessment adjustments and those who do not	Section 5.9
• The processes and systems for calculating and reporting subject results are well-defined and transparent	Section 5.9
• Publicly available QCAA guidelines detail how grade reporting standards are set	Section 5.9
• The approach for calculating general subject results uses the Rasch item response theory (IRT) model, a well-regarded, technically sound method for analysing educational assessment data	Section 5.9
• Use of the Rasch model creates comparable scores both between years and across Internal and External Assessment marks used to calculate subject results	Section 5.9
• A staged approach reduces the complexity of concurrent analysis and the likelihood of error, improving efficiency and accuracy	Section 5.9
Monitoring, review and system impact	
• The governance structure with the QCAA Board and its Standing Committees works well in monitoring and improving the QCE system	Section 6.1
• Implementation of the substantial changes with the new QCE has progressed well with adjustments being implemented where monitoring has already revealed opportunities for improvement	Section 6.2
• The design of data and information processes in the new QCE system is straightforward with support available when needed	Section 6.3
• Information provided to teachers is relevant to teachers' professional practice and students' learning	Section 6.3
• Some teachers perceive that those who do not participate in Confirmation, Endorsement and external marking training do not have access to the same information as those who do	Section 6.3
• Some external stakeholders suggest that data could be provided more directly and efficiently for their organisational needs	Section 6.3

Finding	Report Section
• Public confidence has improved since the new system was introduced	Section 6.4
• Confidence differs by stakeholder groups: school leaders, parents and employers generally indicated confidence in the system whereas school staff were more moderate	Section 6.4
• The new QCE system has guidelines and resources to help students understand academic integrity	Section 6.5
• AI tools, the employment of tutors and plagiarism are major risks to academic integrity	Section 6.5
• QCAA has invested significant resources in increasing expertise within the agency	Section 6.6
• A moderate level of concern was expressed regarding support for QCAA internal staff development	Section 6.6
• QCAA should seek to solidify the new system into a 'business as usual' mode after extensive systemic changes	Section 6.6
• QCAA staff felt empowered to enact the new QCE system	Section 6.6
• QCAA's resources and professional development are viewed positively by school staff members, QCAA staff and school leaders	Section 6.6
• QCAA has developed a suite of assessment literacy training modules that have been widely utilised by Queensland educators	Section 6.7
• QCAA has authorised external reviews of the modules and implemented updates based on the findings to keep content current	Section 6.7
• QCAA has developed and delivered a set of assessment literacy course modules that are available to Queensland teachers	Section 6.7
• Positive views of teacher assessment literacy were found across a range of stakeholders and program participants	Section 6.7
• Most QCE assessments demonstrate strong psychometric properties, with score reliability consistently above acceptable standards (0.83–0.96) and good model fit across most items and criteria	Section 6.8
• Assessment criteria function consistently across years with minimal differential item functioning, though some subjects show other evidence of marking drift over time	Section 6.8
• DIF analysis revealed minimal bias across demographic characteristics (gender, ATSI status, LBOTE, school sector, geolocation, and AARA), providing evidence of the fairness of the assessments for diverse student populations	Section 6.8
• Several subjects show suboptimal targeting of the IAs at the higher achievement levels, creating ceiling and floor effects that limit discrimination among students at the extremes of the performance range and reduce measurement precision	Section 6.8
• Some ISMGs have structural limitations impacting reliability, particularly single criteria carrying excessive weight (e.g., 15-mark criteria) and rating scales exceeding optimal cognitive load for markers	Section 6.8
• External Assessment marking shows acceptable agreement after resolution processes, though some subjects demonstrate initial disagreement levels that warrant further investigation with respect to marker training	Section 6.8
• Moderate to high correlations between IA and EA ability estimates confirm both components measure related achievement constructs	Section 6.8
• Endorsement and Confirmation processes do not significantly affect IA criteria performance. A very small percentage of students experience significant score changes during Confirmation	Section 6.8
• Grade boundary-setting processes are generally sound, with expert judgements mostly consistent with statistical evidence, though misalignment exists for some subjects at key boundaries	Section 6.8

Summary of recommendations

The following summary of recommendations is organised into six broad strategic areas for improvement of the QCE system along with their relevant lower-level recommendations that are presented throughout the body of the report.

Recommendation	Report Section
1. Refine documentation related to the QCE and build a stronger public facing profile of the processes that underpin the elements of the qualification for stakeholders and the wider community	
a. QCAA to develop documentation materials that explain the QCE system's rationale and methodology to diverse audiences across the state to build greater confidence in the system. This includes targeted resources ranging from brief, accessible overviews for parents, students, and community members to detailed methodological guides for education professionals	Section 3.1
b. Given the state's geographic and demographic diversity, prioritise accessible formats and distribution methods that reach remote and regional communities, ensuring equitable access to information regardless of location	Section 3.1
c. QCAA to clearly define the EA's intended purpose and the attributes of quality assessment that EA will address on the publicly available QCAA website	Section 5.4
d. QCAA to consider reviewing its documentation suite to ensure clear content, accurate alignment with enacted operational procedures and sufficient detail to support implementation	Section 5.9
e. That QCAA work with stakeholder groups to identify which data and information are needed and how to communicate that data efficiently	Section 6.3
2. Enhance governance and decision-making through increased external representation and specialist advisory capability	
a. The Queensland Government consider the inclusion of more external members on the QCAA Board, without necessarily increasing the size of the board	Section 6.1
b. The QCAA Board establish a Research Advisory Committee with a majority of external members, including a smaller, specialised group to deal with technical issues	Section 6.8
3. Establish a coordinated research and evaluation agenda to inform continuous improvement, sustainability, and confidence in the QCE system	
a. QCAA to establish a research agenda to identify factors that might require amendment to current syllabuses and then monitor the implementation of changes, including the impact on timeline and workload	Section 4.3
b. QCAA to review how much of the diversity of QCE offerings at the state level can be provided to students within schools and determine whether any limitations in capacity at the school level disadvantage particular groups of students	Section 4.4
c. QCAA to update and revise the assessment literacy training modules to align with contemporary research	Section 5.6
d. QCAA to undertake review of incorporating advanced technology into the QCE assessment processes, including considering the value for money/return on investment and complexity of change management associated with the implementation of such technology	Section 5.8
e. QCAA to commission a study of the user experience, code architecture, accessibility and functionality of the ICT systems supporting the QCE system from a suitably qualified technology team	Section 5.9
f. QCAA consider investigating the extent to which different stakeholders find the new QCE system transparent and how that relates to their perceptions of the fairness of and their confidence in the new QCE	Section 6.4

Recommendation	Report Section
4. Continue to build and sustain system-wide assessment capability to support consistent, high-quality implementation of the QCE across Queensland	
a. QCAA to support further building of capability in schools across the state, particularly for those schools that face challenges due to geographic constraints and teacher shortage	Section 3.2
b. QCAA to continue to build assessment literacy and capacity with new teachers to sustain the level of knowledge across the education sector	Section 3.2
c. QCAA to design and implement a system for providing quality assurance data and feedback to all markers, and ensure that Lead and Chief Assessors meet the pre-determined minimum feedback expectations	Section 5.4
d. QCAA to collaborate with school authorities to encourage all Queensland senior secondary teachers to successfully complete the core assessment literacy training modules within 2 years of employment	Section 5.6
e. That QCAA seek to reduce the information gap between those who participate in Confirmation, Endorsement and External Assessment marking and those who do not	Section 6.3
f. QCAA to implement enhanced marker training and calibration for EA subjects showing persistent elevated initial disagreement levels, with ongoing monitoring and rapid intervention protocols	Section 6.8
5. Continue to strengthen monitoring and review processes to identify risks, unintended consequences, and system impacts across the QCE	
a. QCAA to continue partnering with QTAC to review and monitor any unintended consequences of the reforms, recognising the close relationship between the QCE and tertiary admission	Section 3.1
b. That QCAA prioritise a review of the implementation of the Essential English (2025) curriculum and syllabus	Section 4.3
c. QCAA to review the credit awarded for VET Certificate and Diploma courses in comparison to subjects with syllabuses developed by QCAA	Section 4.4
d. QCAA to incorporate checks to assure adherence to EA development process as designed and monitor for schedule slippage	Section 5.4
e. QCAA to review the marking schedule and evaluate the feasibility and schedule impact of a longer marking period to ensure that sufficient time is allowed for planned training, operational activity and quality assurance tasks	Section 5.4
f. QCAA to consider ways to strengthen quality assurance procedures for Applied subjects	Section 5.5
g. QCAA to clarify rationale and purpose of differential weighting of Internal and External Assessments and select principled and consistent approach across subjects	Section 5.7
h. QCAA to collect and analyse data on casual assessor recruitment and retention, with attention to trends over time that may indicate potential threats to sustainability	Section 5.8
i. QCAA to consider modifying casual workforce schedules to optimise flexibility and improve efficiencies	Section 5.8
j. QCAA to investigate the extent to which experiences with the AARA process vary by different student or school factors (e.g., reasons for AARA, school location or school type)	Section 5.9
k. QCAA to establish a common measurement scale across years to enable monitoring of assessment targeting and student progress over time	Section 6.8
l. QCAA to enhance the psychometric monitoring procedures integrated within the existing quality assurance processes, including enhanced linking approaches and differential item/test functioning analyses, particularly to monitor and address potential marking drift and the functioning of assessment criteria over time	Section 6.8
m. QCAA to develop comprehensive documentation of boundary-setting processes and implement periodic external reviews by independent measurement experts	Section 6.8

Recommendation	Report Section
6. Use longitudinal data analysis and psychometric evidence to drive continuous improvement and quality assurance across QCE assessment and reporting processes	
a. QCAA to analyse Endorsement rates annually within and across subjects for each validity and accessibility priority and examine the results for informative patterns for variables of interest such as geographic location and content area	Section 5.2
b. QCAA to analyse Confirmation rates within and across subjects for each validity and accessibility priority and examine the results for informative patterns	Section 5.3
c. QCAA to consider updating reporting conventions to add more transparency to the number of submissions, approvals and students engaged in the AARA process	Section 5.9
d. QCAA to consider monitoring changes in the distribution of results in Internal and External Assessments over time as an added check on the integrity of Internal Assessments not provided under supervision	Section 6.5
e. QCAA to investigate alignment between assessment difficulty and student ability to maximise precision and eliminate ceiling and floor effects	Section 6.8
f. QCAA to conduct a systematic review of ISMGs using psychometric evidence to revise under-performing criteria, particularly single-criterion assessments and excessively large rating scales	Section 6.8
g. QCAA to combine statistical evidence (equating, predictive models, sequential validation) with expert judgement when setting grade boundaries, following international best practice	Section 6.8

Acronyms

ACARA	The Australian Curriculum, Assessment and Reporting Authority (ACARA) is the independent statutory authority responsible for the development of a national curriculum, a national assessment program, and a national data collection and reporting program that supports learning for Australian students.
ACER	The Australian Council for Educational Research (ACER) is an independent, not-for-profit organisation established in 1930, headquartered in Camberwell, Victoria. ACER is dedicated to advancing educational research and improving learning outcomes across all stages of life, both in Australia and internationally.
AERA, APA, & NCME	American Educational Research Association, American Psychological Association, & National Council on Measurement in Education.
ATAR	The Australian Tertiary Admission Rank (ATAR) is a number between 0.00 and 99.95 in intervals of 0.05, that indicates a student's position relative to all the students in their age group for the purposes of tertiary entrance selection.
ASAT	The Australian Scholastic Aptitude Test (ASAT) was a 3-hour multiple-choice test of scholastic aptitude commissioned by the Australian Council for Educational Research (ACER). At the end of 1974, it was administered in all Australian states and territories except Victoria
ASQA	Australian Skills Quality Authority (ASQA) is the national regulator for Australia's vocational education and training sector.
CERI	The Centre for Educational Research and Innovation (CERI) is an integral part of the OECD's Directorate for Education and Skills. Established in 1967, CERI aims to support member countries in achieving high-quality lifelong learning for all, thereby contributing to personal development, sustainable economic growth, and social cohesion.
FP	Field Position (FP) was a rank order of students within specific subject fields or areas of study, based on their performance in senior secondary subjects for tertiary entrance purposes.
ICT	Information and Communication Technology (ICT) broadly refers to technologies that provide access to information and facilitate communication.
OAI	The overall achievement indicator (OAI) is a fine-grained measure that is divided into bands to form OPs. It is the weighted mean of a student's best 20 semester units of scaled SAIs, including at least three subjects studied for 4 semesters.
OECD	The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation with 38 member countries, founded in 1961 to stimulate economic progress and world trade.
OP	Overall Position (OP) The OP was a tertiary entrance rank order of students based upon their overall academic achievement in senior secondary subjects for tertiary entrance purposes used in Queensland. OP was represented on a scale from 1 to 25 with 1 being the highest.
PISA	The Programme for International Student Assessment (PISA) is a global study conducted by the Organisation for Economic Co-operation and Development (OECD). It evaluates the knowledge and skills of 15-year-old students in reading, mathematics, and science, focusing on their ability to apply this knowledge to real-life situations.

QCAA	The Queensland Curriculum and Assessment Authority (QCAA) is a statutory body of the Queensland Government responsible for kindergarten guideline and senior secondary syllabus development, and for providing resources and services to help teachers develop curriculum, teaching and learning programs from kindergarten to Year 12. It also provides testing, assessment, moderation, certification and vocational education and training services to Queensland's education community.
QCE	The Queensland Certificate of Education (QCE) is Queensland's senior secondary schooling qualification. It is internationally recognised and provides evidence of senior schooling achievements.
QCS	The Queensland Core Skills Test (QCS), commonly referred to as the QCS Test or simply QCS, was a statewide test formerly completed by all Queensland year twelve students who were eligible for an Overall Position, and was optional (but beneficial) for tertiary entrance rank students.
QCIA	The Queensland Certificate of Individual Achievement (QCIA) is awarded by the Queensland Curriculum and Assessment Authority (QCAA). It recognises the achievements of students who have completed at least 12 years of education and have participated in individualised learning programs tailored to their specific needs.
QHRC	Queensland Human Rights Commission
QTAC	The Queensland Tertiary Admissions Centre (QTAC) is a non-profit organisation owned and operated by Queensland universities that manages applications for tertiary study across 17 institutions in Queensland.
SAI	A subject achievement indicator (SAI) is a representation of a student's achievement in a subject. For large subject-groups the SAI is a number — from 200 to 400 —that is assigned by schools to students reflecting their relative achievement in that subject-group. The 200 is the lowest and the 400 is the highest and for each subject-group at least one student must get 200 and one must get 400. For small and intermediate subject-groups the SAI reflects the level of achievement and the rung position within that level. For use in OP and FP calculations, SAIs are transformed to scaled SAIs. SAIs are only given to OP-eligible and equivalent-OP-eligible students.
TES/TEA	In Australia, a Tertiary Entrance Score (or Tertiary Entrance Aggregate, TEA) is a measure of a student's academic achievement in their senior secondary school years, typically used for university admissions. The TEA is a precursor to the Australian Tertiary Admission Rank (ATAR), which is a percentile ranking used by universities for course selection.
UNCRC	United Nations Convention on the Rights of the Child
UNESCO	The United Nations Educational, Scientific and Cultural Organisation (UNESCO) is a specialised agency of the United Nations with the aim of promoting world peace and security through international cooperation in education, arts, sciences, and culture.

Contents

List of Figures

LIST OF FIGURES (CONT.)

List of Tables

01
Introduction

1. Introduction

Queensland has led significant reforms in Australian senior secondary curriculum and assessment over the past 50 years. It abolished external examinations from 1973 and strengthened and refined school-based assessments through a succession of reforms that gave schools substantial autonomy over curriculum and shifted from norm- to criterion-referenced assessment.

A new Queensland Certificate of Education (QCE) system commenced for students entering Year 11 in 2019 and Year 12 in 2020. The new QCE system is the subject of this evaluation.

This change was introduced in response to the recommendations from an independent review of secondary assessment and tertiary entrance conducted by the Australian Council for Educational Research (ACER) (Matters & Masters, 2014). The Queensland Government adopted these recommendations, with minor modifications, in 2016. The new QCE is the culmination of five years of planning and development and reflects a major cultural change in schools, the Queensland Curriculum and Assessment Authority (QCAA), and the wider community.

The new senior secondary assessment and tertiary entrance systems were intended to:

- provide a new approach to senior secondary assessment that combined school-based assessment with External Assessment set and marked by QCAA
- strengthen the quality and comparability of school-based assessments
- provide an Australian Tertiary Admission Rank (ATAR) for students seeking tertiary entrance instead of the previous Overall Position (OP) rank, with responsibility transitioned to the Queensland Tertiary Admissions Centre (QTAC)

Key assessment elements of the policy response (Queensland Government, undated) included:

- new quality assurance processes to strengthen the quality and comparability of school-based assessment
- senior subject results based on four assessments (three school-based and one external for each general subject)
- the introduction of subject-based External Assessment to contribute 25% of the total subject result in most senior subjects and 50% in mathematics and science subjects
- school-based assessment results not scaled against the External Assessment result
- subject results provided as an overall numerical score and a level of achievement (A–E) aligned to syllabus standards

The QCAA was responsible for developing the new system, supported by $109 million in Queensland Government funding. A ministerial taskforce with representatives from the schooling sectors, teacher unions, parent organisations, principal associations, the tertiary sector, the QCAA and QTAC guided the reforms.

The curriculum and assessment reforms were intended to deliver the following high-level outcomes:

- seamless transition — schools, students, tertiary entrance providers and other stakeholders to be well supported through the changes, effectively acquiring a deep understanding of the new system
- world class curriculum —contemporary learning options providing students with 21st-century skills such as ICT, communication, collaboration, and creative and critical thinking, to prepare them for work or study after school
- valid and reliable assessment — strengthened school-based assessment combined with new External Assessment to deliver fair and equitable achievement outcomes for all students
- support for improved teaching and learning — high-quality resources and professional development programs, along with reduced summative assessment, to support teachers and students to make the most of the teaching and learning process
- continuous improvement — advances in educational knowledge and digital technologies embraced to prepare students for the modern economy and benefit employers and universities

The defining characteristics of the new system include:

- syllabuses that describe what must be taught and how student performance must be assessed
- an assessment program of school and QCAA-devised assessments that accumulates evidence of student achievement throughout the senior secondary years
- predefined standards for judging the quality of student achievement
- teacher professionalism where participation in curriculum development and the assessment design and judgement-making process is intended to reinforce pedagogical practice and assessment literacy credentials that provide students with a 'passport' for future pathways that summarises their learning achievements

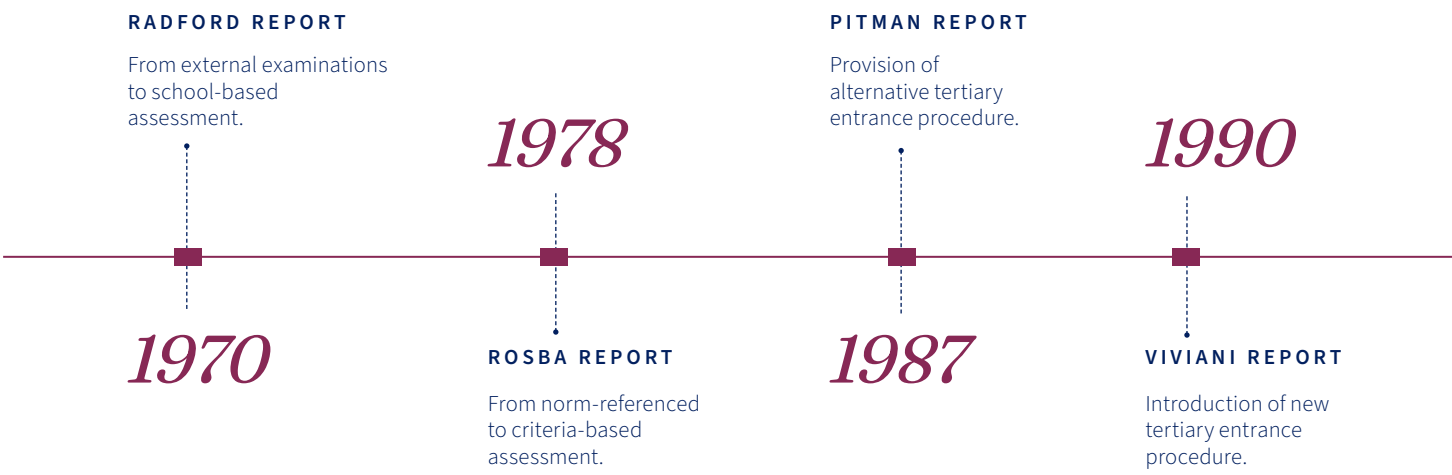


Figure 1.1 Timeline for major changes in Queensland senior secondary education

1.1 Evaluation of the reform

The QCAA commissioned the Assessment & Evaluation Research Centre (AERC) in the University of Melbourne Faculty of Education to undertake an independent evaluation of the reforms. The evaluation focuses on the new QCE system – the policies, processes, procedures and relationships that together provide the foundation for Queensland students' senior secondary schooling experience.

The purpose of the evaluation is to assess the development and implementation of the new QCE system and to make recommendations for future refinements and enhancements. The terms of reference outlined seven elements of the system, to be assessed through 24 areas for review, later reduced to 20 by mutual agreement (see Appendix A):

1. rationale and foundation concepts
2. development and transition processes
3. curriculum
4. new assessment processes
5. supporting policies, procedures and systems
6. ongoing monitoring and review
7. system impact

QCAA required an inclusive approach to the evaluation, with appropriate consultation with the authority, relevant education stakeholders and the wider community.

1.2 History of Queensland's reforms in senior secondary education

The new QCE system followed a series of reforms in senior secondary education that began in 1973 with the implementation of the recommendations of a committee appointed to review the system of public examinations for Queensland secondary schools, chaired by William C. Radford, then Director of ACER (1970). The new QCE is based on recommendations of an independent review of senior secondary assessment and tertiary entrance by Matters and Masters (2014), with Masters then being the Chief Executive Officer of ACER and a successor to Radford.

Between these reviews, others also amended and developed the structure and governance of senior secondary education in Queensland (see Figure 1.1). These developments reflected changes in demographics and the purposes of senior secondary schooling, leading to changes in curriculum and assessment in the final two years of schooling and in selection for post-school destinations.

1.2.1 Changes in demography and purposes of senior secondary education

The Radford Committee (1970) noted that, while the percentage of students progressing from Year 8 to Year 12 had slowly increased — from 26.3% for students in Year 8 in 1964 to 28.8% for students in Year 8 in 1969 (p. 77) — the percentage of those completing Year 12 who progressed to full-time tertiary education declined, as shown in Table 1.1 (p. 78).

The proportion of students progressing beyond Year 12 to full-time university study dropped from 34.2% in 1961 to 24.2% in 1968. Nevertheless, the upper secondary curriculum and end-of-secondary assessments were heavily influenced by preparation and selection for university study. The Radford committee concluded that while the Year 12 examinations had been "intended to select for university education … secondary education up to the Grade 12 level now has more purposes than that of preparing students for study at university" (p. 76).

Table 1.1 Percentage of Year 12 students by following year destinations

Year	Percentage of students progressing from Year 12 to:			Percentage not progressing to further study
	University	Teachers Colleges	Other advanced education	
1961	34.2	21.3		44.5
1962	35.7	23.9		40.4
1963	33.2	22.2		44.6
1964	30.0	18.4		51.6
1965	26.2	14.7	2.6	56.5
1966	28.4	16.6	2.1	52.9
1967	27.2	16.3	4.0	52.5
1968	24.2	16.8	6.7	52.3

These two trends continued to the point where Matters and Masters (2014, p. 2) observed that participation of 17-year-olds in Year 12 had increased from 31.8% in 1977 to 75.7% in 2013, while the percentage of students in Year 12 studying subjects that provided eligibility for tertiary entrance declined from nearly 100% in 1977 to less than 60% in 2013.

In the meantime, the New Deal Report (Pitman et al., 2002) outlined a vision of enabling all young people to complete a post-initial stage of education, to be recognised by the award of a Senior Certificate if they achieved "at least a threshold quantum of Worthwhile Learning" (p. 7). All learning was to be under the auspices of a school but could occur in a range of settings. The QCE was introduced in 2008.

A review of the QCE after its initial operation concluded that it was "widely accepted and appreciated in Queensland … with its broad conceptualising of learning combined with externally moderated school-based assessment, the QCE is a strong model for student learning in the 21st century" (Hunter, 2011, p. 68). The review recommended improved communication about the value and utility of the QCE, exploration of ways to facilitate early banking of credits in student learning accounts (so that the account becomes a working reference point as students progress through the senior phase of learning), and the development of new procedures to prevent students "double dipping" in accumulating credit for learning (pp. 69–71).

1.2.2 Changes in curriculum

In response to demographic change in the late 1960s, the Radford Committee (1970) proposed broadening of the curriculum offered in Years 11–12. Subject syllabuses would still be provided by the Board of Senior Secondary School Studies, but schools would be given "freedom in the choice of detail of content [and] in the method of presentation" (p. 80). The report also envisaged new subjects being developed by schools and others, including subject associations, that could be approved by the Board for trial and, if found to be suitable, be included among Board subjects. In addition, the report proposed that subjects be offered in half-yearly (semester) units

instead of as two-year subjects to allow students to achieve greater breadth of study (p. 81).

The Review of School-Based Assessment (ROSBA) Report (Board of Senior Secondary School Studies, 1978) was primarily focused on assessment. In his preface, Chair Professor Edward Scott referred to achieving a "higher retention rate of students beyond Grade 10 [and] increased awareness of the multicultural mix of the secondary school population" (p. ii). The report maintained the subject breadth introduced following the Radford Report but introduced new procedures for review and accreditation of subjects with state panels (p. 22). Specifically in relation to curriculum, the review recommended that the three categories of subjects be retained but that more use be made of Board-approved subjects and school subjects to broaden student choice and increase school autonomy (p. 21).

The reports of the Working Party on Tertiary Entrance (1987) and the Viviani review of tertiary entrance in Queensland (1990) focused mainly on procedures for selection of students for tertiary study, but both addressed curriculum issues.

The Working Party report observed that the Board of Secondary School Studies, in "striving to fulfil the intentions of the Radford committee to give both teachers and students greater flexibility and freedom in curriculum matters and to promote the view that secondary education should be seen as a stage that has its own objectives," had introduced a new category of Board-registered school subjects. These were developed and offered by a school on its own initiative, with results reported on Board certificates. Results in some TAFE subjects were also approved for inclusion in Board certificates (Board of Senior Secondary School Studies, 1978, p. 46). The report acknowledged that "so long as results in senior secondary school studies contribute information to the selection process [for tertiary education] there will be backwash effects on the secondary school curriculum", but added that the issue was not about whether there would be backwash effects but, rather, what those effects should be (p. 100).

Viviani (1990) noted that the Radford Report "broke the direct control of the University of Queensland over the secondary curriculum [and] led to the increasing autonomy of schools in the development of curriculum and assessment, the widening of course offerings and the professional upgrading of teachers" (p. 1). The report added, "this review supports the values that underpin that ongoing process of reform and seeks to strengthen these" (p. 1). However, it acknowledged, as had the Working Party report, that the "TE [tertiary entrance] 'culture' dominates our upper secondary schools, in its effect on curriculum and assessment" (p. 19).

1.2.3 Changes in assessment

Abolition of external examinations in Queensland

Queensland abolished external examinations at the end of Year 12 after 1972, judging that the harmful effects outweighed the benefits. The benefits included:

- students being motivated to work towards a tangible goal and an objective record of achievement
- Teachers being able to compare student achievements across schools
- Schools receiving information on the quality of their teaching

The harmful effects included (Radford Committee, 1970, p. 78):

- students being rewarded for reproducing others' ideas and for short-lived knowledge
- teachers being pressured to do work better left to students, being constrained by the syllabus, and having to adjust to examiners' idiosyncrasies
- schools focusing less on important outcomes not examined externally

On balance, the committee recommended replacing external examinations with school assessments. Where an order of merit was required, it should be based on school assessments scaled against External Assessments not based on prescribed syllabuses (p. 3).

Results in the Senior Examination were primarily normative, creating a comparative ranking of students to support tertiary selection. Subject results in the external examinations were expressed on a 1–7 scale, with 7 representing the highest achievement. Prescriptions guided the percentage of grades awarded: 2–6% awarded grade 7, 6–12% grade 6 and so on. Eligibility for enrolment at the University of Queensland required students to reach a minimum total score based on their best five subjects.

Introduction of school-based assessments

While some school-based assessments had been introduced alongside external examinations in other jurisdictions, the Committee's recommendation to remove external examinations and rely exclusively on school-based assessments for subject results was seen as an adventurous move. The Australian Capital Territory was the only other Australian jurisdiction to follow suit (McGaw, 1977).

The shift led to successive changes over time to strengthen the quality, validity and reliability of those assessments. Initially, the aim was to build public confidence that assessments were comparable across schools. Later, reforms sought to transform assessments from normative to standards- or criterion-referenced assessments while still ensuring they fulfilled a normative function: ranking students for post-school destinations where demand exceeded supply. The changes are well documented by Clarke (1987, 1990).

From 1972, school-based assessments were conducted at the end of each of the four semesters in Years 11 and 12. This contrasted with external examinations, which had only been held at the end of Year 12. Because students were now assessed four times in each subject, the minimum aggregate required for admission to the University of Queensland was set at four times the minimum that had applied in 1971 under the external examination system.

A moderation process was established to ensure comparability across schools. Chief Moderators were appointed for ten subject areas. They were supported by District Moderators, who conducted teacher meetings to review samples of student work submitted by schools and establish consensus on standards and grade allocations (Clarke, 1987, p. 21).

Although assessments were expected to align with prescribed grade distributions, allocations soon departed from the model. The Board attributed this partly to pressure to promote university admission, with some schools perceiving themselves as advocates in moderation meetings to secure more high grades for their students (Clarke, 1987, p. 27).

Other factors also contributed. After the first semester, grade distributions showed more high grades and fewer low grades, as some Year 11 students with low grades in Semester 1 had dropped subjects. In subjects such as French and German, which had selective enrolments of higher achieving students, there was reluctance to limit the number of top grades, as the model required. These factors effectively resulted in rescaling to reflect student cohort characteristics, contributing to grade inflation. As a result, the University of Queensland raised the minimum aggregate requirement significantly, with the prospect of further escalation if grade inflation continued.

To relieve pressure on school-based assessments, an alternative tertiary selection method was introduced. At the end of Year 12, schools were required to provide a single final assessment on a 1–99 scale for each student in each Board subject. The scores were then rescaled to match mean results on the Australian Scholastic Aptitude Test (ASAT) and set with a standard deviation of 12. An aggregate was obtained by summing the best adjusted scores for subjects studied across 20 semester units, with at least three subjects studied over four semesters. Each student's initial aggregate was calculated as a weighted sum of adjusted scores for each subject, weighted by the number of semester units studied. The aggregate scores for all students in the school were then rescaled to the ASAT mean and standard deviation for the cohort. The results became the students' Tertiary Entrance Scores (TES) (McGaw & Hannan, 1985, p. 71).

Evidence of the Aptitude Test's utility came from correlations between ASAT scores and school-based assessments between schools and within schools. A lower between-schools correlation, based on school means, indicated rescaling was necessary for comparability, while a higher within-schools correlation supported the validity of ASAT rescaling (McGaw et al., 1975). An earlier study had shown "an order of merit based on school assessment scaled against ASAT would be as effective as that based on aggregate scores in a student's best five Senior Examination subjects for the award of [Commonwealth University] Scholarships" (Research and Curriculum Branch, 1972).

Other institutions used the TES in 1974, but the University of Queensland did not until 1975 (Clarke, 1987, p. 28). This reduced pressure on moderated school-based assessments, which no longer determined tertiary selection but instead served to record student achievement. Moderated school assessments remained competitive norm-referenced results, with students categorised into grades according to prescribed distributions, even with some deviations. These and subsequent changes sought to balance three to some extent competing demands:

- certifying student achievement in each subject studied
- taking account of differences in the general abilities of candidates in different subjects
- creating an overall order of merit for selection for post-school destinations for which demand exceeds supply of places

Review of school-based assessments

In the years that followed, important changes were made to strengthen and alter the role of school-based assessments. A report from a special committee established by the Board and chaired by Professor Ted Scott recommended abandoning the normative distribution of grades in favour of performance-based criteria. The report argued that it was student achievement that should be assessed, "rather than how they compared with other students" (Board of Senior Secondary School Studies, 1978).

The Board determined that subject results would be expressed on a scale 'highest level of competency', 'highly competent', 'competent', 'limited competence', 'very limited competence', and 'ungraded' (Clarke, 1987, p. 33). For Years 11–12, only one assessment at the end of Year 12 would be published, rather than one for each semester (Board of Senior Secondary School Studies, 2011, p. 19). This emphasised competency at the end of the study period. It also avoided the issue of differences in the general abilities of candidates in different subjects by expressing achievement against subject-based criteria.

Moderation of school-based assessments was phased out over a three-year period — in 1983 for 17 schools in Brisbane and Townsville, in 1984 for schools in the Brisbane–South Coast, Townsville and Cairns districts and in 1985 for the remaining schools. To move from norm-referenced to criteria-referenced school-based assessments subject committees redeveloped syllabuses that stated objectives as student outcomes in terms of cognitive skills, content knowledge and affective objectives (attitudes, values and feelings). The committees also provided criteria for assessing students' levels of achievement of the objectives and suggested assessment techniques to help maintain standards and comparability between schools in the absence of moderation (Clarke, 1990, pp. 1–4).

As the changes were introduced, complaints arose about excessive administrative requirements for teachers and schools, and excessive testing of students to judge their achievements. Some schools and teachers also struggled to adopt assessment based on outcomes and performance criteria (Clarke, 1990, pp. 5–7). To assist schools, the Board established an Assessment Unit that operated from 1985 to the end of 1987. The Unit produced 21 discussion papers on a range of topics, published from 1986 to 1988 (Board of Senior Secondary School Studies, 2011). An ongoing difficulty in public debate was confusion

between the school-based criteria-referenced assessment, used for reporting subject achievement, and the norm-referenced TES, used for tertiary selection (Clarke, 1990, p. 10).

Review of Tertiary Entrance Score system

After its introduction in 1974, the TES remained largely unchanged until the 1980s. At the same time, public discussion ranged from general acceptance by the higher education institutions, to criticisms of the ASAT, to calls for the TES to be abandoned in favour of school-based assessments (Clarke, 1990, pp. 13–14). The latter proposal would require school-based criteria-referenced assessments to serve a norm-referenced function as well.

In 1983, a Working Party on Tertiary Entrance was established by the Minister's Joint Advisory Committee on Post-Secondary Education and the Board of Secondary School Studies, chaired by John Pitman, Executive Officer of the Board. The Working Party's report (1987) made 52 recommendations for change, including the introduction of an Overall Achievement Position (OAP) Profile and subject-specific Field Positions to provide additional, relevant information for tertiary faculties. The review also recommended adding a writing task to the ASAT to create a new Common Scaling Test (Clarke, 1990, pp. 17–19).

The recommendations of the Pitman Committee were not supported by "the University of Queensland and some institutions other than universities and, in a period of ministerial change, were not acted upon. It was in these circumstances that the newly elected Goss government, faced with an impasse in previous policy making on the issue and in a situation of increasing public disquiet, abolished the Tertiary Entrance Score with effect from 1992" and commissioned a new inquiry with Professor Nancy Viviani as Chair (Viviani, 1990, p. 2).

The Viviani Committee's terms of reference included avoiding the use of a single score to measure students' aptitude for tertiary study and avoiding the rescaling of school-based assessments. The Committee recommended a profile similar to that proposed by the Working Party, with students receiving an Overall Position (OP) and Field Positions (FPs) in four areas:

- extended written expression
- short written communication
- basic numeracy
- complex problem solving with mathematical symbols and abstractions

The FP scores were to be created from weighted results in relevant subjects. Universities were expected to give greatest weight to the OP, and to use the FPs as supplementary information. The report also recommended continuation of scaling with the Core Skills Test but proposed improved moderation procedures that might replace scaling within subjects but not between subjects (Viviani, 1990, pp. 39–47).

Reintroduction of external subject-based assessments

Matters and Masters (2014) recommended that external examinations be reintroduced. However, their proposal did not represent a simple return to the practices abandoned in the 1970s. While adding externally assessed information for post-secondary selection, it preserved a strong component of school-based assessments.

Consultations by Matters and Masters revealed support for the flexibility and professionalism of teacher assessment, but also concerns about the quality of some teacher-devised assessments and the adequacy of moderation processes intended to ensure comparability across schools (pp. 23–24).

They recommended discontinuing OPs, FPs and the Queensland Core Skills Test. Instead, they proposed subject results that would indicate a student's level of knowledge, understanding and skill, based on one External Assessment set and marked by QCAA, and three school-based assessments marked on a 10-point scale according to QCAA specifications. School-based assessments would not be scaled against External Assessment results. The three school-based assessments would together carry the same weight as the single examination result.

Moderation of the school-based assessments would involve:

- *endorsement* by Assessment Supervisors that the proposed assessment activities fit the QCAA specifications
- *confirmation* through blind re-marking of samples of students' work by teachers from other schools to assure comparability of assessments across schools
- *ratification* at the end of Year 12 to confirm that there are no anomalies in a school's results

The Queensland Government's response to these recommendations instituted a major reform of senior secondary curriculum and assessment, commencing with students entering Year 11 in 2019 (Queensland Government, undated). The announcement stated:

> Teachers will continue to exercise their professional judgement in designing and administering school-based senior assessment. Senior students will continue to be provided with multiple opportunities to demonstrate their skills and knowledge. Under the new arrangements, these features will also be complemented by subject-based External Assessment and new processes to support and promote high-quality assessment practice. (Queensland Government, undated, p. 1).

This change was characterised as combining "the flexibility and authenticity of school-based assessment, developed and marked by classroom teachers, with the additional rigour of a common External Assessment set and marked by the QCAA" (Queensland Government, undated, p. 2).

There would be three school-based and one External Assessment in each subject. The External Assessment would contribute 25% of the result in most subjects, and 50% in mathematics and science subjects, on the grounds that "the type of skills and knowledge found in these subjects are more readily assessed through common External Assessment" (Queensland Government, undated, p. 1).

There would be no scaling of school-based results against the External Assessments, and no scaling of subject results to take account of differences in subject candidature. However, scaling would occur in the creation of an order of merit for tertiary selection, where subject results were adjusted to match the distribution of a students' results in other subjects.

The overall rank would be expressed as an ATAR — a percentile rank from 99.95 downwards — indicating a student's position relative to all students in the age group (assuming those not receiving an ATAR rank below those who do). QTAC is responsible for the creation of the ATARs from the assessments provided by QCAA.

The government adopted the recommendations with modifications, committing to a reduction in the number of assessments, strengthened school-based assessments, the introduction of External Assessments, and a more inclusive tertiary entrance rank requiring satisfactory completion of one of the five QCAA English subjects (Queensland Government, undated).

1.2.4 Changes in governance

Prior to the implementation of the Radford Report, the senior secondary curriculum and the Year 12 examinations were managed by the University of Queensland. The Radford Report (1970), recommended the creation of a single Board of Secondary School Studies to manage the curriculum and assessment for the lower secondary (Years 8–10) and upper secondary (Years 11–12) (p. 5). The Board was established in 1971, shifting responsibility for the senior secondary curriculum and assessment from higher education to secondary education.

From 1989 to 2002, governance of senior secondary education was narrowed through the creation of a Board of Senior Secondary School Studies (BSSSS). At the same time, a Tertiary Entrance Procedures Authority (TEPA) was established. TEPA formally separated responsibility for curriculum and assessment in senior secondary education from the management of tertiary selection, which was shared between the BSSSS and the QTAC.

From 2002 to 2014, the BSSSS and other curriculum and assessment bodies were merged into the Queensland Studies Authority (QSA). In 2014, the QSA was recast as the QCAA. QCAA is governed by a seven-member board comprising nominees from the state, Catholic and independent school sectors, as well as four ministerial nominees selected on the basis of expertise relevant to the functions of the authority (QCAA, 2025d). Ministerial nominees typically include some from the tertiary education sector.

1.3 Changing purposes of senior secondary of education

The purposes of senior secondary education in Australia have evolved significantly. A growing proportion of students now continue to upper secondary education while university education has become less dominant as a post-school destination. A senior secondary certificate of education (SSCE), such as the QCE, is designed to provide students with a foundation for transition to a range of post-secondary pathways, including work, further study and participation in civic life. All states and territories offer such a certificate as part of the Australian Qualifications Framework (AQF).

Recent assessments, such the report of the Review of Senior Secondary Pathways into Work, Further Education, and Training (Shergold et al., 2020), highlight the need for jurisdictions to continue developing their SSCEs to meet changing community demands. Key recommendations include developing a common understanding of digital literacy across senior curriculum, and an agreed set of capabilities to equip students with essential life attributes, including literacy, numeracy, employability skills, creative and entrepreneurial capabilities, financial literacy, interpersonal skills and civic understanding. These reviews routinely add to the criteria against which the QCE is judged.

Despite the broadening role of senior secondary education, tension persists at the secondary-tertiary education interface. The QCE defines and reports against standards of learning in subjects. Reducing this information to a single rank order in the ATAR inevitably creates tension, although the growing use of alternative pathways to tertiary education may ease this pressure.

The QCE must continue to adapt to meet these evolving purposes.

1.4 System context

The SSCE reflects the communities and contexts within which it is developed and operates. As the community contexts become more complex, so too do the influences upon SSCEs. The QCE operates in a policy and practice environment shaped by many factors, including:

- changing technologies including the influence of large language models, and generative artificial intelligence (AI)
- increasing demand for greater choice and more flexible approaches to teaching and learning in senior schooling
- the status of the teaching profession, the professional identity of teachers and workforce supply pressures
- participation and retention issues such as diversity in STEM, First Nations outcomes, equity, declining arts enrolments nationally and the role of vocational education and training (VET) in schools

- contemporary issues such as globalisation, climate change, information disorder, the increased availability of AI and the role of schooling systems and curriculum programs in responding to these

The QCE is delivered to nearly 55,000 students annually across more than 500 state, independent and Catholic schools. A notable feature of the Queensland context is the geographic and economic diversity of the communities served. While there is significant concentration of population in south-east Queensland, the QCE is also delivered in large regional centres, rural communities and remote locations.

When the new QCE system was being developed and implemented, education sectors in Queensland (and elsewhere) were facing significant challenges. These challenges affected reform delivery and are therefore relevant to this evaluation:

- Teacher workforce shortages – these shortages were more acute in regional and remote locations, with greater proportions of early career teachers and teachers teaching out-of-field. This was a challenge for a qualification like the QCE, which relies on teachers with high levels of curriculum and assessment literacy and appropriate discipline knowledge.
- COVID-19 pandemic disruptions – the pandemic emerged in early 2020, coinciding with the first cohort of students completing the new QCE. Like all jurisdictions, Queensland schools had to respond to rapidly changing health directives. The impact was substantial. In 2020, students completed only two Internal Assessments instead of three. In 2021, public health restrictions required QCAA to regularly modify its approaches. These factors complicated an already demanding change-management activity.
- 2022 flooding disruptions – widespread flooding in south-east Queensland disrupted schools and the QCAA itself. QCAA offices were closed at short notice, requiring staff to implement disaster recovery measures and remote working protocols

It is notable that the first cohort of students to graduate from the new *QCE* system in 2020 did so in the first year of a global pandemic.

1.5 Authorising environment

The QCE system and QCAA operate within a complex authorising environment of diverse stakeholder groups and legislative requirements. QCAA's functions and powers are set out in the *Education (Queensland Curriculum and Assessment Authority) Act 2014* (Qld) ('the QCAA Act') and *Education (Queensland Curriculum and Assessment Authority) Regulation* 2014 (Qld). The Act details QCAA's responsibilities in syllabus development, testing, moderation and certification.

QCAA does not own or operate schools. Queensland school operations fall under several legislative instruments, including the *Education (General Provisions) Act 2006* (Qld), the *Education (Queensland College of Teachers) Act 2005* (Qld), and the *Education (Accreditation of Non-State Schools) Act 2017* (Qld). Implementation of curriculum and assessment programs rest with schools and governing bodies. QCAA supports schools and provides quality oversight of the certificates.

A core function in the QCAA Act is syllabus development. This includes the requirement for QCAA to "adopt the Australian curriculum content and standards of achievement for the subject as the basis for developing the syllabus" (Section 9(2)) where a senior secondary Australian curriculum exists. Unlike other jurisdictions, QCAA does not develop syllabuses for P–10, as Queensland adopts the P–10 Australian Curriculum as written.

Schools or learning providers must open a student account within one year before a student enters the post-compulsory participation phase (where the student stops being of compulsory school age — less than 16 years or after completing Year 10). In practice, students can commence their QCE once they turn 15 or start Year 10.

QCAA issues the QCE, the Queensland Certificate of Individual Achievement (QCIA) and the statements of results. It is not responsible for calculating the ATAR. This evaluation considers the QCE system in relation to senior secondary certification only and does not include factors associated with calculation and administration of the ATAR, which is managed by QTAC. Impacts and consequences of the introduction of the ATAR in Queensland are outside the scope of this evaluation.

1.6 Structure of this report

This report evaluates the foundation of the new QCE system and its introduction (Chapter 3), changes in curriculum (Chapter 4), changes in assessment (Chapter 5), with a detailed analysis of the psychometric properties of the new assessments (Chapter 9) and provision for ongoing monitoring and review (Chapter 6). Conclusions are presented in Chapter 7.

02
Methodology

2. Methodology

2.1 Evaluation Design

This section describes the mixed methods approach undertaken for the evaluation of the QCE.

The evaluation of the QCE system employed a comprehensive, multi-method approach to data collection, with stakeholder consultations occurring from June 2023 to September 2024. Rigorous analytical techniques were employed for the quantitative data, including psychometric modelling of achievement scores. The research team met with QCAA throughout the phases of the evaluation to fact-check initial understanding of several documents and data provided. Dialogue with QCAA staff throughout the evaluation period facilitated clarifications and ensured contextually informed interpretations while maintaining methodological independence.

2.2 Data collection and analysis methods

The evaluation has been conducted in stages. In the first stage, document analysis and a literature review informed the evaluation framework, which was developed in consultation with QCAA in 2023 (Aston et al., 2024). From this work, five overarching evaluation questions were established and addressed throughout the evaluation to address the following areas:

01

To what extent is the new QCE system design evidence-informed, fit-for-purpose, and responsive to contextual changes? (Design)

02

What is the degree and fidelity of implementation across all contexts? (Implementation)

03

To what extent has the new QCE system achieved its intended goals? (Effectiveness)

04

How sustainable is the new QCE system? (Sustainability)

05

To what extent has the investment in the QCAA and the new QCE system provided value for money? (Value for money)

In 2025 the University of Melbourne and QCAA agreed not to pursue Question 5 following meetings on the interim report.

In the second stage, consultations and qualitative data collection were undertaken during visits to 10 schools in Queensland in 2024, as illustrated in Figure 2.1. These consultations provided contextual information that informed the quantitative analysis and the writing of the final report in the third and final stage in 2025.

The eight key research methods employed throughout the evaluation are presented in Figure 2.1 and described in Section 2.3 below. A description of how each research method yields relevant data for the evaluation is provided in Section 2.3.

Ethical approval was obtained from the University of Melbourne Human Research Ethics Committee. Data collection, storage and presentation were conducted in accordance with ethical standards.

Figure 2.1 Consultations and data collection activities

1 **Online data collection activity undertaken during QCAA School Leader Forums held at:**

a. Emerald
b. Hervey Bay
c. Toowoomba
d. Cairns
e. Townsville
f. Mackay
g. Rockhampton
h. Sunshine Coast
i. Gold Coast
j. Brisbane

2 **10 school case studies involving interviews/focus groups with principal, principal's delegate, heads of department, teachers, students, parents & community members. School selection criteria include:**

- At least 4 State schools and at least 2 schools from each of the Independent and Catholic sectors
- Schools from all four location types (major cities, inner regional, outer regional, remote/very remote)
- One school from Far North Queensland
- A range of school sizes, types and indexes of community socio-educational advantage
- At least one school with a high proportion of Aboriginal or Torres Straight Islander students
- At least one school with a high proportion of Language Background other than English students

3 **General survey**

Open to school staff, past QCE students, parents/carers, employers and community members

4 **Targeted surveys**

For endorsers, confirmers, external assessment markers & invigilators, common internal assessment moderators, professional development participants, QCAA staff

5 **Written submissions**

Open to any individual or organisation within Queensland

6 **Observations, focus groups & interviews**

Observations of assessment events

Focus groups and interviews with QCAA board & staff, QTAC, teacher unions, principal and parent associations

7 **Document analysis**

Review of all relevant QCAA project and process documentation, policies, guidelines and reports

8 **Analysis of existing QCAA data (deidentified)**

- Internal assessment data
- External assessment data
- ICT development and usage data
- Student enrolment and achievement data
- Human resource and finance data
- Professional development attendance and satisfaction data
- Communication data

2.3 Data collection and participants

Data collection and consultations with different stakeholders from June 2023 to September 2024 were organised through the following methods, capturing diverse perspectives and data:

1. Ten *school leader forums* were conducted to gain perspectives from school leaders in a range of schools and contexts. School leaders are key stakeholders in the QCE system, responsible for the operationalisation of the intended design and directly aware of any barriers to high-fidelity implementation. Data were collected during the forums using an online tool and results were discussed with participants.
2. Ten *school case studies* were undertaken to examine how the new QCE system has been implemented within school contexts, including capturing perspectives from members of the schools' wider communities. The ten schools were selected by the University of Melbourne, based upon criteria including location (urban, suburban, rural or remote) and school type (government, independent or Catholic). These consultations were organised by QCAA and provided insights into the new system from student groups, teachers, schoolteachers, parents and representatives from local community industries. (Details of the school statistics are provided in Table 2.1 below.)
3. *A general survey* was conducted to gather broader perspectives on the new QCE system from multiple stakeholder groups such as school staff, past QCE students, parents/carers, and other interested participants (community members, employers, etc.). In addition to the selected-response items, respondents were invited to provide free-response comments to a range of questions about their perceptions of the system. These comments were analysed for themes that are described in the text of the report.
4. Additionally, *targeted surveys* were conducted with specific groups including endorsers, confirmers, QCAA staff, invigilators and External Assessment markers. The assessor surveys were designed to gather data from the QCAA's casual workforce performing roles related to implementation of the system. All targeted surveys offered respondents the opportunity to provide written comments and submissions to gather more detailed qualitative information (see Appendix E). The comments and submission were analysed for themes that are described in the text of the report.
5. *Observations* were conducted across multiple QCE assessment activities including Endorsement, Confirmation, applied subject quality assurance, specialised training for new Assessors and External Assessment processes. These observations were designed to provide detailed insights of QCE assessment activities run by the QCAA. In addition, QCAA provided access to training and workshop materials online, which has been used to better understand the evaluation context.
6. *Focus groups and interviews* were conducted with key stakeholder groups including past and present chairs of the QCAA Board, QCAA staff and external stakeholders. These were designed to gather key stakeholder group perspectives and to provide deeper insights into areas of interest flagged from the analysis of observations, surveys and written submissions.
7. *Document analysis* was conducted throughout the evaluation – first for the evaluation framework, and later for the evaluation of data collected to answer relevant evaluation questions and to benchmark documents against best practice. Documentation reviewed included materials from the extensive QCAA website, internal QCAA documents provided as part of the evaluation and materials within the secure QCAA portal.
8. *Psychometric data analysis* was conducted to establish the psychometric properties of a sample of nine General subjects selected by University of Melbourne. This included Rasch modelling to evaluate the Internal and External Assessment components, produce reliability estimates, evaluate item fit and gauge score targeting. Differential item functioning analysis was conducted to check for bias by gender, Aboriginal and Torres Strait Islander status, language background other than English, location, and access arrangements and reasonable adjustments (AARA), where sample size was sufficient. Analysis of grade boundaries was also conducted.

Table 2.1 Demographic, Socio-Economic Status (SES) and geographic information of the ten school case studies

School	ICSEA*	Sector	Year 12 cohort	% Year 12 Indigenous	% Year 12 LBOTE*
School A	1073	I	Medium	4.03	16.10
School B	1001	C	Small	6.25	2.08
School C	935	G	Medium	9.23	1.54
School D	1038	G	Large	3.03	15.48
School E	1163	I	Medium	0	35.71
School F	825	G	Medium	56.9	41.38
School G	993	C	Medium	2.04	0
School H	1015	I	Small	11.76	2.94
School I	924	G	Large	10.14	9.68
School J	1113	G	Large	0.21	49.27

* ICSEA - Index of Community Socio-Educational Advantage; LBOTE - Language background other than English

The evaluation process was guided by the Evaluation Framework (Aston et al., 2024). Throughout the year, research staff engaged in ongoing discussions to interpret qualitative data in line with the evaluation's terms of reference. The analysis of qualitative data was systematically documented in evaluation summary sheets, which played a key role in facilitating the triangulation and integration of findings across multiple data sources. This approach ensured that interpretations were robust, consistent and supported by evidence from diverse perspectives.

Supported by the QCAA, a large number of participants contributed to the evaluation, providing data from diverse roles and perspectives. Table 2.2 shows the total number of participants, organised by data collection method, and details their involvement across different research activities. This highlights the breadth of stakeholder engagement and the diversity of viewpoints informing the evaluation findings.

Table 2.2 Types of data, research method, and total participants involved in the evaluation

Types of data	Research method	Total Participants
1. School Leader Forum Data collection	**Online survey**	**343**
2. Ten school case studies	**Online / Offline Interviews**	**278**
- *Principals*		*12*
- *Head of Departments and non-principal school leaders*		*57*
- *Teachers*		*98*
- *Current students*		*59*
- *Past students*		*9*
- *Parents and community*		*43*
3. General survey	**Online survey**	**2,682**
- *School staff*		*839*
- *Past students*		*341*
- *Parent*		*135*
- *Other*		*26*
- *Open response*		*1,341*
4a. Targeted surveys	**Online survey**	**2,141**
- *Endorsers*		*313*
- *Confirmers*		*599*
- *Invigilators*		*271*
- *Markers*		*887*
- *QCAA staff*		*71*
4b. Written submissions	**Online questionnaire**	**52**
5. Observation	**Online / Offline workshop observation[1]**	**N/A[1]**
- *Endorsement[2]*		*N/A*
- *Confirmation[3]*		*N/A*
- *External assessment[4]*		*N/A*
- *New Assessor training[5]*		*N/A*
6. Focus group interviews	**Online / Offline Interviews**	**97**
- *External stakeholders*		*35*
- *Remote beginning teachers*		*5*
- *Remote school leaders*		*4*
- *Remote employers*		*3*
- *Remote local council*		*3*
- *Remote indigenous community representatives*		*4*
- *Remote past students*		*5*
- *QCAA staff*		*36*
- *QCAA Board chairs*		*2*

Types of data	Research method	Total Participants
7. Document analysis	**Throughout the evaluation**	N/A
8. Psychometric data analysis	**Subject-specific analyses**	9 General subjects (see Appendix B).

[1] Data not available on how many participants were involved in the online/offline endorsement/confirmation/external assessments/assessor training workshops.

[2] UoM academics observed three pre-event meetings; the endorsement events took place in February 2024.

[3] UoM academics observed one Confirmation App demonstration, one Confirmation Decision Panel (CDP) meeting, one general new Assessor training workshop, and three subject-specific new Assessor training workshops from April to June 2024.

[4] UoM academics observed two subject-specific grade boundary meetings, two script selection sessions, three external assessment events and six marker training sessions from October to November 2024.

[5] UoM academics observed one online training session for new Assessors led by QCAA staff in August 2025.

2.4 Strengths and limitations

The mixed-methods approach to evaluation has inherent strengths in supporting triangulation of quantitative findings enriched with context from the qualitative results. This specific evaluation was limited by the broad scope. It is never possible to gather all relevant data and viewpoints when conducting a review of sample of stakeholders and viewpoints. This evaluation had the support of the QCAA in gathering information from a very broad base of stakeholders, so the span of the work is as comprehensive as possible.

Some areas for review required specialised skills for in-depth analysis of the topic, such as econometric analysis of value for money and detailed software functionality analysis of the IT systems and custom applications. These areas for review were removed from this evaluation by mutual agreement between the QCAA and University of Melbourne. Recommendations have been included to retain evaluators with the specialised expertise necessary to complete a considered and comprehensive review of these areas.

03

Foundation concepts and system development

3. Foundation concepts and system development

3.1 Rationale and foundation concepts

Area for review 1.1

The accessibility and quality of information on the rationale for the introduction of the new system, its theoretical and methodological underpinnings, and its essential policies and procedures.

This chapter presents evaluation findings for the first two elements of the new QCE system identified in the Terms of Reference:

Successful implementation of education reforms depends on having a clear vision and purpose, communicated in accessible language, and made available for teachers and schools, in a timely way, regardless of school location, size or resources (Burns et al., 2016). Research also shows that successful implementation involves sharing information with other stakeholders, including parents, researchers, journalists, unions, policymakers, and the broader public (Hopfenbeck, 2016; Hopfenbeck et al., 2013). Material and policies need to be distributed through multiple channels: online, through professional development and in print when needed. Content should be clear, free from jargon and appropriate for the target audience. New procedures are usually supported with model examples (for instance, sample assessment procedures) and professional learning before implementation. A clear vision and purpose, well-defined assessment methods and comprehensive policy and procedure manuals that explain system operations are therefore essential.

One example of transparency in education is the International Baccalaureate (IB), which publishes detailed documents[1] that explain the foundational principles behind its diploma system. The IB provides resources that describe the philosophy, vision and purpose of the qualification and the associated assessment practices, along with detailed policy and procedure manuals. The QCE and QCIA Handbook explains the reasoning behind the QCE system and provides background information. Schools can access the handbook through the QCAA portal, direct distribution to principals and professional development sessions. A limitation, however, is that the handbook is written primarily for teachers and education professionals, which may make it difficult for parents, students and the wider community to understand and use effectively.

QCAA's professional development materials clearly articulated the link between the government's policy response to the review recommendations and the resulting system changes. New assessment policies outlined in the handbook build on recent research (DeLuca et al., 2016), and detailed guidelines for teachers and school leaders provide further explanation (QCE & QCIA Handbook, 2025).

1 at https://www.ibchem.com/root_pdf/Principles.pdf and https://www.ibo.org/contentassets/1cdf850e366447e99b5a862aab622883/assessment-principles-and-practices-2018-en.pdf

Data analysis revealed that, while online information was accessible, some teachers reported varied levels of understanding of the methodological foundations, particularly assessment processes. This appears to reflect a need for teachers to build confidence in their own understanding through experience, rather than a lack of quality information.

Regional schools reported greater difficulty accessing in-person professional development sessions that explained the rationale in depth. Queensland has long-standing strategies for supporting remote areas in the state, and feedback suggests this should remain a priority.

Findings
• The QCE and QCIA Handbook successfully provides clear, high-quality information about why the new system was introduced and explains its key policies and procedures. The accessibility and quality of policy procedure information appear to meet the needs appropriately for teachers and school leaders, who are the target audience for these policies.
• The current documentation may not be sufficiently accessible for some parents, students and the broader community who also need to understand the system. More detailed information about how assessment processes are designed and developed could benefit some stakeholders. This would help clarify the underlying goals and educational philosophy that drive the QCE system and its supporting processes.
Recommendations
• QCAA to develop documentation materials that explain the QCE system's rationale and methodology to diverse audiences across the state to build greater confidence in the system. This includes targeted resources ranging from brief, accessible overviews for parents, students, and community members to detailed methodological guides for education professionals
• Given the state's geographic and demographic diversity, prioritise accessible formats and distribution methods that reach remote and regional communities, ensuring equitable access to information regardless of location

Area for review 1.2

How the objectives of the ACER review have been met in the government policy position and the new system.

The ACER review (Matters & Masters, 2014) made 23 recommendations about Queensland's senior secondary assessment and tertiary entrance systems.

The Queensland Government's policy response to the ACER recommendations, Advancing Futures, guided QCAA's implementation of the QCE system. The status of the recommendations is evaluated as follows:

- **Delivered**: The government policy position reflects the recommendation, and QCAA has implemented it directly into operational delivery.
- **Delivered in part**: Some elements of the recommendation have been accepted, but not all.
- **Alternative approach delivered**: The recommendation has been modified or adapted while maintaining its overall intent.
- **Not delivered**: No evidence that was found that the government acted upon the recommendation.

Table 3.1 summarises the status of each recommendation. Where the recommendation falls outside the scope of this evaluation (e.g., tertiary entrance-related recommendations), they are not listed.

Table 3.1 Status of senior secondary assessment recommendations of the ACER review

Recommendation	Status
Recommendation 1 The Overall Position (OP) system should be discontinued and the interface between secondary completion and university selection should be redesigned. The implications are that Subject Achievement Indicators (SAIs) would no longer be generated, Overall Achievement Indicators (OAIs), OPs and Field Positions (FPs) would no longer be calculated, and the Queensland Core Skills (QCS) Test would be discontinued. Under the new model, subject results would be reported on a finer scale for use by universities in their selection decisions.	**Delivered in part** The OP was discontinued and subject results are reported on a finer scale.
Recommendation 2 Responsibility for certifying student attainment in senior subjects should be separated from responsibility for selecting applicants for admission to university courses. The former should be the responsibility of QCAA, working directly with schools. The latter should be the responsibility of the universities and their agent QTAC.	**Delivered**
Recommendation 3 Student attainment in each Authority Subject should be reported by QCAA in the form of a subject result indicating the level of knowledge, understanding and skill that the student has attained. Subject results should be directly comparable across teachers and schools, and function as stand-alone measures of senior secondary attainment, independently of how they might subsequently be used.	**Delivered** A detailed set of specifications across quantity, quality, pattern, and literacy and numeracy requirements for student to obtain credit in core courses of study from applied, applied (essential), general, general (extension) and general (senior external examination) subjects, VET and IBDP in Queensland schools has been delivered. The contribution of credits towards a QCE or QCIA is also clearly defined. Comparability is assured through quality assurance throughout learning delivery, including moderation processes within schools; Endorsement and Confirmation of assessments aligned to standardised curricula for Internal Assessments; and administration and marking of externally set assessments.
Recommendation 4 The certification of student attainment in each senior subject should be based on a set of four specified types of assessment activities. QCAA should specify the nature of each activity, the conditions under which it is to be completed and the marking scheme for assessing students' performances. One of the four assessment activities should be externally set and marked by QCAA.	**Delivered** A comprehensive operational instantiation of the four specified types of assessments has been built, providing detailed and comprehensive syllabuses and Instrument-Specific Marking Guides (ISMG) for each of the three required Internal Assessments.

Recommendation	Status
Recommendation 5 Students' subject results should be reported as integers on a scale of 1–60. Each subject result should be calculated as the sum of a student's mark on the External Assessment (in the range 0–30) and marks on the three assessment activities set and marked by teachers (each in the range 0–10). Teachers' assessments should not be statistically scaled against the External Assessment.	**Delivered** The sum of student marks for the IA and EA determines the subject result as recommended. Teachers' assessments are not statistically scaled against the External Assessment. **Alternative approach delivered** The final policy position has not constrained the mark totals to the 1–60 scale as recommended. The total number of marks for all general subjects is set out of 100.
Recommendation 6 An External Assessment in each subject should be set and marked by QCAA and completed at the same time under the same supervised conditions in all schools. If resourcing is an issue, priority should be given to developing External Assessments for subjects with high enrolments, subjects which are foundational for university courses and subjects for which External Assessment is most practicable. For most senior subjects, the External Assessment should contribute 50% of the subject result.	**Delivered** External Assessments in all general, general (extension) and general (SEE) subjects are set and marked by QCAA, with a few exceptions for small enrolment, mostly language subjects that are shared by other jurisdictions. **Alternative approach delivered** The Ministerial Taskforce established to consider the government response to the report determined that the External Assessment contribute 50% of the final subject result for mathematics and science subjects. In other subjects, it contributes 25%.
Recommendation 7 Three school assessments should be specified for each subject. The nature, intentions and parameters for these three assessment activities should be specified by QCAA, with teachers in schools annually designing local versions of each. The three school assessments and the External Assessment should be designed jointly to provide appropriate coverage and balance of the subject syllabus and in general should address different kinds of learning and achievement within the subject.	**Delivered** General subjects have three Internal Assessments set and marked by the schools in Units 3–4. The course and assessment contents, structure and activities are specified in the syllabuses and ISMGs and balanced across the set of assessments. **Alternative approach delivered** Applied subjects have four Internal Assessments set and marked by schools, in accordance with specifications, conditions and response requirements set by QCAA. In the applied subjects, Essential English and Essential Mathematics, one of the four assessments is a common Internal Assessment that is set by QCAA but marked by the schools.

Recommendation	Status
Recommendation 8 QCAA should assure the validity and reliability of school assessments in each subject through a revised approach to moderation that includes three elements: Endorsement, Confirmation and Ratification. • Endorsement of proposed assessment activities – For each of the three school assessments, QCAA checks locally-devised assessment activities/ instruments and marking schemes for their consistency with QCAA specifications and endorses their use with students. • Confirmation of accurate application of marking schemes. For each of the three school assessments, QCAA checks that schools' applications of marking schemes are accurate and consistent across teachers and schools. This is done through moderation meetings where teachers undertake blind re-assessments of student work against the relevant 10-point scale. QCAA also conducts annual spot sampling and blind re-assessments to check the consistency of marking across schools. Where a problem is identified, all student work in that subject in that school is re-marked. QCAA will determine which assessments in which subjects in which schools will have moderation meetings in a particular year. • Ratification of subject results – At the end of Year 12, QCAA checks each school's results on the four assessments for anomalies. If anomalies are identified, these are investigated and resolved before verifying students' marks on the four assessments. Once resolved, the ratification of students' subject results for certification follows. An appeals process will be available to students after they receive their Senior Statements from QCAA (or in some other way as determined by QCAA). QTAC should be included in discussions about the appeals process.	**Delivered: Endorsement** Templates, ISMGs and technology support for development of the three required Internal Assessments by schools are provided by QCAA. Assessors with no conflict of interest evaluate the provisional assessment instruments for accuracy and consistency with the designated syllabus, in a system with provision for requests for additional information and provision of additional support as needed. Endorsed instruments are used with students in classrooms. **Alternative approach delivered: Confirmation** The accuracy of the schools' application of the ISMG is checked by QCAA through a sampling procedure. Schools are expected to complete internal moderation prior to submission of samples for Confirmation. Samples are specified by QCAA based on provisional marks provided by the school. Assessors with no conflict of interest evaluate the provisional marks for accuracy and consistency, and the system includes provision for requests for additional information and provision of additional support as needed. Approaches to improve quality include requiring additional samples, re-marking of student work and adjustment of marks to compensate for a cohort pattern. **Alternative approach delivered: Ratification** A two-stage ratification process is undertaken. School results are checked for anomalies if a disruption occurs, and results are adjusted prior to being given to students. In the next stage, the Ratification Committee seeks to identify and understand any large differences between Internal and External Assessments for subject cohorts. The QCAA will work with these schools in the following year to understand the discrepancies and address any concerns. **Delivered: Appeals Process** Students and schools have procedures to request review of Confirmation decisions, External Assessment results and information on statements of results at several QCAA-specified points in the assessment cycle.
Recommendation 9 QCAA should establish a guild of Assessment Supervisors to lead the proposed moderation processes (the Endorsement of assessment activities, the Confirmation of the accurate and consistent application of marking schemes, and the ratification of subject results), and to assist in teacher capacity building.	**Alternative approach delivered** Although there is not a guild of assessment supervisors *per se* established by QCAA, assessors (Endorsement) and assessors (Confirmation) generally continue in service in their roles and can move up the hierarchy from assessor to Lead Assessor to Chief Assessor as their period of service extends, given individual interest and demonstrated quality of work. Thorough training (online and in-person) for each role is offered and includes the opportunity to serve as an alternate in the leadership role of interest.

Recommendation	Status
Recommendation 10 The Senior External Examinations (SEEs) currently developed by QCAA should be discontinued. Instead, all students who are undertaking an authority subject should be required to complete the four assessment activities specified by QCAA for that subject (the three school assessments and one External Assessment).	Alternative approach delivered The Senior External Examinations (SEEs) have largely been discontinued. A small number continue to be offered, predominantly for small-enrolment languages. This phasing out of the SEEs has occurred in parallel with greater focus and clarity around the flexible alternative learning options for students wishing to achieve a QCE or enter tertiary study.
Recommendation 15 The Queensland Government should make the legislative changes required to divest the QCAA of its current responsibilities relating to tertiary selection (including scaling and aggregating results to produce rank orders of tertiary applicants).	Delivered
Recommendation 16 The Queensland Government should invest additional funding in creating high-quality assessment and certification processes to underpin a reformed senior secondary credential. A priority order of subjects should be established if it is not possible to fund the development of externally set and marked assessments in all senior subjects.	Delivered The Queensland government has invested in creating the new QCE system, defining new processes and standards, creating detailed manuals for all roles, standardising procedures and roles, building an extensive online presence, developing custom software applications to support assessors and staff and offering extensive training and resources for schools and teachers who implement the new system.
Recommendation 17 QCAA should continue to build its staff capacity in educational assessment, educational measurement and information and communication technologies.	Delivered Training of existing staff and onboarding new staff with expertise in the areas necessary to support schools and teachers as they implement the new curriculum, syllabuses and assessment procedures and use new technology to support quality assurance and assessment procedures has occurred. Ongoing professional development is offered as evidenced in the annual professional learning budget for staff.
Recommendation 18 QCAA should include in its specified assessments processes a greater focus on skills and attributes now being identified in senior secondary curricula as essential to life and work in the 21st century (for example, teamwork, problem solving, creativity, verbal communication).	Alternative approach delivered While is it apparent from documentation around the development and delivery of syllabuses that QCAA has identified and defined a set of 21st century skills as underpinning factors in the senior syllabuses along with literacy and numeracy, QCAA's assessment systems do not yet incorporate an explicit focus on routine assessment of each of these skills. Some skills are assessed in some subjects (e.g., English and Languages assess verbal communication and Technologies subjects assess problem solving)
Recommendation 19 The Queensland Government should devise a multi-platform information strategy to precede and accompany any significant changes or reforms to senior assessment and tertiary entrance.	Delivered The Advancing Futures policy initiative (Queensland Government, undated) was supported with an accompanying communication strategy.

Recommendation	Status
Recommendation 23 As part of the Queensland Government's commitment to further development of the Queensland Certificate of Education, consideration should be given to enhancing the capacity of QCAA to develop and deliver a world-class senior secondary qualification. This may include establishing a group of specialist staff within QCAA capable of further conceptualising, leading and implementing the recommendations of this Review. It may also include the creation of two separate authorities, one with responsibility for curriculum and assessment in Years K–9, the other with responsibility for Years 10–12.	Delivered in part QCAA has invested in training of existing staff as well as onboarding new staff with expertise in the areas necessary to continue the evolution of the QCE and QCIA as world-class senior secondary qualifications. QCAA has increased internal expertise and added external advisory capacity to support envisioning the future state of the system. QCAA retains responsibility for K–12 curriculum and assessment in Queensland.

Findings

- The recommendations of the Matters and Masters review (2014) have largely been adopted either in full or part in the Queensland Government policy position and have subsequently been implemented by QCAA

Recommendations

- QCAA to continue partnering with QTAC to review and monitor any unintended consequences of the reforms, recognising the close relationship between the QCE and tertiary admission

3.2 Development and transition processes

Area for review 2.1

The success of the development process, including the various assessment trials and other testing activities that informed the development of current processes and procedures, and the effectiveness of the QCAA's program and project management.

Successful development processes when implementing education reforms are known for being grounded in extensive consultations, have a phased implementation, and build clear communication as well as capacity building and use feedback mechanisms throughout the process. This allows adjustments to be made throughout the implementation of the new reforms (OECD, 2016) and trust between stakeholders (Hopfenbeck et al., 2013).

As documented in this section, QCAA conducted extensive consultations with subject experts from schools and universities, had multiple trials over three years, provided detailed guidance for schools and provided teacher education and professional development. Feedback mechanisms were included to ensure QCAA could evaluate and improve the processes based upon the experiences throughout the implementation process and adjust the implementation based upon feedback received from stakeholders.

When implementing the new QCE in 2019, the intention was to modernise Queensland's assessment approach while aligning it more closely to national standards. The reform, officially titled *Advancing Futures: New Senior Assessment and Tertiary Entrance Systems in Queensland* (Queensland Government, undated), employed a systematic and evidenced based approach to the development of the new QCE system, which included a number of Confirmation, Endorsement and External Assessment trials in a range of subjects. These gave schools and students the opportunity to familiarise themselves with the new system and to evaluate the effectiveness of the respective processes. The trials assessments were developed collaboratively with subject experts from schools and universities and administered under secure conditions. The trials were furthered implemented in several phases, with the following subjects: Semester 1, 2016: Chemistry, English, Geography, Mathematics B, Modern History; Semester 2, 2016: Economics, Graphics; Semester 1, 2017: English, French, Japanese, Mathematics B, Physical Education, Visual Art.

Endorsement and Confirmation trials were conducted parallel to quality assurance activities associated with the former QCE system. These trials served not only to test and develop teacher understanding of the new processes but also strengthened the quality assurance practices for those cohorts completing their QCE in the former system.

External Assessment trials were developed for students in Year 11. The schools that participated in these trials received detailed information about individual student achievement and cohort analysis, allowing them to understand the implications of the new assessment approach without impacting students' Year 12 exit results.

To support the implementation of the reforms, QCAA established a Program Management Office (PMO) and adopted project management methodology. The PMO reported regularly to the SATE Governance Board, which comprised senior QCAA personnel, as well as the Audit & Risk committee and the QCAA Board The Department of Education also established the Senior Secondary Assessment Governance Group that included the Director General of the Department of Education, the three schooling sectors, the QCAA CEO and Chair, and Queensland Treasury and Department of the Premier and Cabinet officers.

Findings

- Consultations with stakeholders provided information that the program was successfully delivered throughout the challenging times during the pandemic, but the perceptions of the success of delivery varied across the ten school cases

Recommendations

- QCAA to support further building of capability in schools across the state, particularly for those schools that face challenges due to geographic constraints and teacher shortage

Area for review 2.2

QCAA's stakeholder communication and change management strategies to ensure that teachers and schools understand the new system and are appropriately skilled to exercise their responsibilities in their school context.

We know from the literature that successful implementation of stakeholder communication and change management strategies involves the stakeholders having a clear knowledge of what the changes entail to ensure that they can create understanding for the changes needed. It further involves building capabilities and capacity so that teachers and school leaders have the skills needed to clarify and establish role clarity. In this project, it involved defining and making clear what was the responsibility of QCAA and what was the responsibility of teachers and school leaders at the school level.

We further know that successful communication and change management strategies involve using consultation processes wherein stakeholders are guided through the transition period and not left unsupported. Productive dialogue between the different levels was therefore crucial for successful involvement of stakeholders throughout implementation phases.

This also involves identifying potential resistance in the change processes and gathering feedback and insights from stakeholders involved to make sure these sources of resistance can be tackled early in the process. The dialogues with stakeholders involved will more likely secure successful implementation, as it is demonstrated that participatory processes increase the commitment and ownership from staff involved (OECD, 2016).

As part of the evaluation, the general survey demonstrated evidence from staff on the following areas: 884 out of 894 respondents (99%) reported in 2024 that they had a good understanding of the new QCE system (Strongly Agree, Agree, or Somewhat Agree). Although the data were collected in 2024, and not all teachers who responded to the survey were teachers at the time of the implementation, the survey results can be seen as an indication of long-term implementation of ideas from the new reform. This can be noted as a successful implementation of the core system, with the reservation that we have a somewhat low response rate for this survey (39%). Nevertheless, the fact that the respondents to this study have reported this positive outcome is an indication of a successful implementation.

Stakeholders reported a strong understanding of the new QCE system and QCAA's responsibilities in implementing the system. Improvement in Endorsement and Confirmation rates suggest teachers have continued to build their understanding over time, and teachers self-report greater confidence in their understanding five years into the reforms. A challenge for QCAA will be to maintain this knowledge and understanding as new teachers enter the system.

Findings

- Survey data demonstrated a strong perceived understanding of the new QCE system among stakeholders. This finding was corroborated through consultation meetings with teachers and school leaders across all ten case studies, triangulating the survey results and confirming their reliability

Recommendations

- QCAA to continue to build assessment literacy and capacity with new teachers to sustain the level of knowledge across the education sector

Area for review 2.3

The organisational change process within the QCAA and whether appropriate systems are in place to ensure its internal workforce continues to develop and attract new expertise, embrace new ideas and ways of working, and become an employer of choice for teachers and other professionals.

Best practices for organisational change and workforce development in educational settings emphasise the importance of evidence-based approaches, strategic communication and structured change processes. In public organisations such as QCAA, the normative commitment to change, such as the obligation and duty to support change, can be a driving force mitigating negative reactions and resistance.

Recent research demonstrates that organisations that have been able to develop into what is known as learning organisations, with continuous professional development, tend to attract employees with higher skills and ambitions to become life-long learners (Mikkonen et al., 2017).

Several key documents, including QCAA's strategic plans and annual reports, document the commitment within the QCAA to develop and sustain a learning organisation, using feedback from different stakeholders to improve. The organisation further emphasises the importance of building cultural capability across the QCAA, equipping the organisations to better support Aboriginal and Torres Strait Islander peoples in their work as part of Queensland's education community. QCAA's recruitment practices demonstrate a commitment to diversity and inclusion, emphasising the inclusive culture of their organisation and the value of having people from culturally diverse backgrounds.

The 2022–23 Annual Report notes a 96% retention rate among the 300 full-time equivalent staff in June 2023. The lockdown and COVID-19 pandemic could potentially be factors in the high retention rate at this point, but it is worth noting. The report further explains how staff are supported by formal performance and development processes and leadership programs. This stability is important to note, with recent reviews in other systems (e.g., VCAA's *Incident Review*; Blacher, 2025) highlighting staff turnover as a major risk for organisations like QCAA.

Findings

- The QCAA demonstrates a strategic commitment to continue being a learning organisation that attracts and retains skilled professionals through continuous professional development, strategic workforce planning and a robust, inclusive organisational culture

04
Curriculum

4. Curriculum

4.1 Curriculum development

Area for review 3.1

The QCAA's effectiveness in developing new syllabuses and resources and providing other support for curriculum development and delivery in schools. This should include consideration of the QCAA's syllabus development, review and revision processes and activities, and whether the new syllabuses meet the objectives set in the development phase (e.g., 21st century skills, best practice in teaching and learning, currency, and responsiveness to advances in knowledge, government initiatives and stakeholder needs).

This chapter presents findings regarding the third and fifth elements of the new QCE system identified in the Terms of Reference:

A high-quality curriculum depends on specific policies in the state/context and country in which it is implemented. The quality of the curriculum comprises both the content of the curriculum and the development process.

4.1.1 Quality curriculum content

Over many years, the OECD's Centre for Educational Research, and Innovation (CERI) has refined the notion of high-quality curriculum through research and consultation with member countries. For example, the 1990 report Curriculum reform: an overview of trends (OECD/CERI, 1990) draws upon case studies submitted by OECD member countries. A major OECD conference in 1993 on the theme Curriculum Redefined examined findings in five areas: (1) Learning to Think, Thinking to Learn; (2) Core Curriculum; (3) Student Assessment and Evaluation; (4) Science, Mathematics and Technology Education; and (5) Humanities, Arts and Values. This work, published as *The Curriculum Redefined: Schooling for the 21st Century* (OECD/CERI, 1994) emphasised providing opportunities for all students by differentiating the curriculum based on individual abilities.

OECD work has also suggested including problem solving, technology literacy, creativity and adaptability, teamwork, recreational activities and interdisciplinary activism while avoiding curriculum overload and supporting diverse programs and teaching methods to meet students' varied talents and interests.

This work influenced international conceptions of curriculum quality and underpinned the progressively developed assessment framework used in the OECD international surveys, Programme for International Student Assessment (PISA) (OECD, 1999). The first PISA assessed students in reading, mathematics, and science, with the measurement focus on how they could use what they know rather than what they have learned (OECD, 2001). Following the first PISA results, some countries incorporated this emphasis in their curricula and assessments, reducing the focus on rote knowledge (Schleicher, 2009). For example, Japan and South Korea, changed their assessments to incorporate open-response items in admission tests for universities (South Korea) as well as in national assessments (Japan).

While PISA continues to assess reading, mathematics and science, its frameworks evolve with contemporary views of curriculum. In PISA 2025, science is again the main domain with assessments of students' ability to engage with science-related issues as reflective citizens and to use scientific ideas for informed decision making, framed as three competencies (OECD, 2023). These competencies — explaining phenomena scientifically, constructing and evaluating designs for scientific enquiry and interpreting scientific data and evidence; and researching, evaluating, and using scientific information for decision-making and action — represent major goals for science education.

Since 2000, PISA has also measured cross-curricula competencies (e.g., problem solving, as analogical reasoning in 2003, collaborative problem solving in 2015, and financial literacy in 2022). PISA 2025 is assessing Learning in the Digital World, which measures students' capacity for iterative knowledge building and problem solving using computational tools. This capacity is demonstrated by effective self-regulated learning while applying computational and scientific inquiry practices. The tasks involve coding and modelling (OECD, 2025a). These developments introduce new emphases within existing subjects rather than adding new subjects to be fitted into a common curriculum in P–10 or as new subjects in the senior secondary years (OECD, 2025b).

OECD's "curriculum for the future" work outlines how curricula have evolved over time. In the 19th century, the focus was preparing students for the job market through static, linear, standardised academic disciplines such as mathematics and languages. In the 20th century, influenced by the World Wars and the Cold War, the focus shifted to preparing students to become independent and personally fulfilled. Curricula broadened in scope but often remained standardised. For the 21st century, the vision is education for citizenship and a balance between breadth and depth. Globally, curricula are now more dynamic, flexible, and personalised (OECD, 2019, p. 12).

In recent years, curriculum development has increasingly incorporated scientific evidence showing that students' active participation is essential for learning. This has supported a shift toward providing students with agency through a student- and learner-centred curriculum (Freeman et al., 2014). This approach challenged traditional curricula in which teachers were seen as the sole owners and transmitters of knowledge. From this perspective, high-quality curricula also need to be flexible so they can support student agency and active learning.

More recent debates have reintroduced direct instruction as a teaching strategy, supported by cognitive load research by Sweller and colleagues at the University of New South Wales. Their work shows that student problem solving with unfamiliar material can create a cognitive load that exceeds students' capacity for learning (e.g., Sweller et al. (2011); Tindall-Ford et al. (2020).

The choice of the instructional mode is a matter for teachers and schools. For curriculum design, the issue is whether the curriculum supports a range of pedagogical approaches. Since 2020, OECD has published thematic reports based on consultations with countries around the world. These reports highlight key factors for successful curriculum development and implementation (OECD, 2020). Consultations with countries such as Belgium, Chile, Denmark, and Estonia demonstrated that a flexible curriculum can enhance teaching effectiveness and inclusivity. Drawing on this evidence, the OECD emphasises collaborative policy making and teacher education as essential strategies. It also notes that clear goals, accountability mechanisms, and societal support are crucial for successful curriculum implementation.

To summarise, OECD's global collaboration identifies a set of features important for future-oriented curricula. Table 4.1 shows these features and the extent to which QCAA is able to influence them in Years 11–12.

Features of future-oriented curriculum	Opportunity for QCAA influence in Years 11–12
1. **Core Foundations**: Basic skills and knowledge needed for all further learning	Limited
2. **Transformative Competencies**: Creating value, handling tensions, taking responsibility	Yes
3. **Agency**: Students' ability to act purposefully and collaboratively	Some
4. **Knowledge**: What students need to understand (facts, concepts, procedures)	Yes
5. **Skills**: What students need to be able to do (cognitive, social, physical)	Yes
6. **Attitudes & Values**: Guiding principles that influence decisions and actions	Yes
7. **AAR Cycle**: Learn through Anticipation, Action and Reflection	Yes
8. **Best Practice in Teaching and Learning**: Responsive to advances in knowledge and government initiatives and stakeholder needs	Some

Table 4.1 OECD recommended features of future-oriented curriculum

A curriculum authority like QCAA has limited influence over students' development of basic literacy and numeracy skills in the final years of schooling (Item 1). Its influence in this area comes mainly though curricula for earlier years of schooling, particularly in primary schools. However, QCAA can influence practices in teaching and learning in senior secondary education where subject curriculum shape pedagogy in schools (Items 3 and 8). For the other items, curriculum content and specifications provide direct opportunities to influence teaching and learning.

4.1.2 Evaluation of QCE curriculum content

A national collaborative effort through ACARA produced a national curriculum for a Foundation Year and Years 1–10 in all subjects, and for Years 11–12, 15 subjects across English (4), Mathematics (4) Science (4), History (2) and Geography (1). The F–10 subjects have been introduced across the country, with minor variations in some States and Territories. Queensland adopted them as written. Implementation of the Years 11–12 subjects depends on State and Territory determination.

The *Education (Queensland Curriculum and Assessment Authority) Act* 2014 (Qld) requires adoption of the content and standards of achievement of Australian Senior Secondary Curriculum national subjects as the basis for developing the detailed syllabuses (Queensland Government, 2014). Queensland's curriculum includes three of the four national English subjects as general subjects and Essential English as an applied subject. It adds a general extension course and a short course in literacy. In mathematics, it includes the four national courses, three as general subjects and Essential Mathematics as an applied subject, and a short course in numeracy. In Science, it includes the four national subjects as general subjects and an additional three applied subjects. In Humanities and Social Sciences, it includes the two History and the Geography national subjects.

All general, general extension and applied subjects offered by QCAA are provided in syllabuses that are tightly structured. Each syllabus includes a course rationale, subject matter, assessment specifications, and an ISMG or Instrument-Specific Standards (ISS). These features are intended to ensure consistency across schools and student groups in delivering subject knowledge, literacy skills and criteria for judging the quality of student work. They apply no matter where students live or what school system, they are in.

Overall, teachers were satisfied with the content of the new syllabuses. Case studies revealed positive feedback in many subject areas, including the general subjects, English, Geography and Digital Solutions, and the applied subjects, Aquatic Practices and Sport and Recreation. In the applied subject Industrial Technology Skills, stakeholders noted better alignment with industry.

The only 2019 syllabus to receive sustained criticism was Essential English. It is now in its final year of implementation and will be replaced by the revised 2025 syllabus for graduating students in 2026. In discussions of the 2019 Essential English syllabus in consultation meetings, staff, employers and past students identified what they *perceived* to be a mismatch between the knowledge and skills it required and those needed for vocational education or work. There was broad agreement across data sources that the subject should place greater emphasis on clear, succinct written communication, a skill needed in many workplaces.

A challenge was noted in aligning the syllabus with the national Essential English curriculum while also progressing students to Level 3 of the Australian Core Skills Framework (ACSF; McLean et al, 2012), the national competency standard for literacy and numeracy. Some students undertaking Essential English have not yet mastered Year 10 expectations, making the leap to ACSF Level 3 difficult. Teachers also reported that Essential English is much more academically rigorous than in the past, which can make it difficult to teach to the wide range of abilities in these classes. While some suggested recalibrating the syllabus to better reflect student needs, reducing its complexity would not meet the ACSF Level 3 requirement. A short course in literacy provides an alternative pathway for students who might not be suited to Essential English. Some teachers also reported a significant gap between 2019 Essential English and General English, making it difficult to meet the needs of students at the upper end of Essential English and the lower end of General English.

Case study reports on applied subjects varied by discipline, but in general participants expressed satisfaction with the new curriculum. The inclusion of applied subjects was noted as a strength of the curriculum. Their inclusion was seen as a positive response to the wide range of post-school pathways that students may be interested in taking and to differences in abilities and interests among students. Stakeholders highlighted the inclusion of applied subjects as a strength, noting their value in catering to students with different abilities, interests and post-school pathways. The possibility of having ATAR based on four general subjects and one applied subject instead of five general subjects only, allowing students to combine general and applied subjects to achieve a QCE, was viewed positively.

The extent to which the Queensland senior secondary curriculum reflects the OECD's recommended features is shown in Table 4.2, with the match level based on a review of QCAA documents. In practice, the match depends on schools' implementation of syllabus content and advice, but QCAA's published syllabuses provide strong links.

Features commended by OECD	Queensland curriculum match to features
1. **Core Foundations**: Basic skills and knowledge needed for all further learning	Clearly provided in subject syllabuses and through short courses in literacy and numeracy
2. **Transformative Competencies**: Creating value, handling tensions, taking responsibility	Clearly provided in complementary (21st century) skills within syllabuses
3. **Agency**: Students' ability to act purposefully and collaboratively	Included in complementary (21st century) skills within syllabuses
4. **Knowledge**: What students need to understand (facts, concepts, procedures)	Clearly defined in syllabuses
5. **Skills**: What students need to be able to do (cognitive, social, physical)	Clearly defined in syllabuses
6. **Attitudes & Values**: Guiding principles that influence decisions and actions	In some syllabuses
7. **AAR Cycle**: Learn through Anticipation, Action and Reflection	Provided, where relevant, in assessments involving student experiments and investigations
8. **Best Practice in Teaching and Learning**: Responsive to advances in knowledge and government initiatives and stakeholder needs	Assistance provided by access online to resources helpful in teaching and creatively assessing students' learning

Table 4.2 Comparison of Queensland curricula with OECD commended features

4.2 Syllabus development

In the senior secondary years, Queensland students can follow either a Queensland Certificate of Education (QCE) pathway or a Queensland Certificate of Individual Achievement (QCIA) pathway. The QCE pathway can include general subjects (including general extension subjects available only as Unit 3 and 4 subjects), applied subjects, and vocational education and training (VET) certificate courses, provided they do not duplicate learning in applied subjects already studied. The QCIA pathway is available for students with a disability "that affects learning and is not primarily due to socioeconomic, cultural and/or linguistic factors" (QCAA, 2025b). QCAA is responsible for general and applied subjects. VET qualifications are the responsibility of registered training organisations, but QCAA determines whether there is duplicated learning in an applied subject.

4.2.1 Quality syllabus development, review, and revision

The key considerations in this section relate to the nature of curriculum and syllabus documentation, the process of their creation, and the program for revision. On documentation, van den Akker et al. (2009) identified 10 key components of quality (see Table 4.3). Curriculum and syllabus documents may provide only a broad framework and leave the development of all detail to schools and teachers. They typically outline aims and objectives and specify content at a broad level, perhaps in terms of the standard of learning outcomes expected. In such cases, the teacher is declared or implied to be central in elaborating content, sourcing resources for the teacher and students, managing teaching and learning logistics (including student grouping, location, and time allocation), and assessing student learning.

Rationale
Aims and objectives
Content
Learning activities
Teacher role
Materials and resources (learning activities)
Grouping
Location
Time
Assessment

Table 4.3 Features of high-quality curriculum & syllabus documentation

According to UNESCO, OECD, and research on syllabus development, the development of high-quality syllabuses requires the features shown in Table 4.4 (Jenkins, 2020; OECD, 2018, 2019, 2020; Stabback, 2016; Voogt et al., 2016).

Planned and systematic
With clear goals
Inclusive and consultative processes
Led by curriculum professionals
Cyclical in nature
Sustainable

Table 4.4 Requirements to develop high-quality syllabuses

Best practice in syllabus review and revision usually involves an ongoing, systematic process in which data are collected from relevant stakeholders through regular consultations. Across different recommended frameworks, there is general agreement that high-quality reviews of syllabuses require committees or panels including subject-matter experts, teachers, leaders, administrators and researchers. These panels interpret and advise on next steps based upon stakeholder feedback and comparisons with research-based standards and exemplary models. Local teacher knowledge helps ensure redevelopment of material is fit for purpose.

Collaboration between experts ensures that learning progressions, assessment tools, and implementation support are grounded in local needs, while also establishing monitoring systems for implementation and future developments. As in formative assessment, where teachers improve their teaching based upon student feedback and performance data, review and revisions of syllabuses require ongoing mechanisms. Teachers and students must inform judgements on how the system is working and how it can be improved (Darling-Hammond, 2009; Wiliam, 2011). This requires feedback loops in which implementation data inform further refinements. This iterative approach ensures that curriculum revisions respond to both immediate needs and to long-term goals, while maintaining alignment with research on effective instructional design.

4.2.2 Evaluation of the QCE syllabus development processes

Current QCAA syllabuses are more prescriptive than the previous Queensland syllabuses. This is intended to reduce ambiguity about what must be taught and assessed. This contrasts with earlier arrangements such as those recommended in the Radford Report (1970): "Syllabuses approved by the Board will present a broad framework of the subject and not prescribe its detail" (p. 2). Similarly, the ROSBA report recommended continuation of this broad framework approach (1978, p. 21). The move to more prescriptive syllabuses aligns with the new assessment procedures recommended by Matters and Masters (2014), although that report did not specifically address syllabuses or course content. (Assessment processes for general and applied subjects are discussed in Chapter 5.)

Nevertheless, the new QCE system maintains flexibility for schools in the delivery. The QCE and QCIA Policy and Procedures Handbook (QCE & QCIA Handbook (2025) makes clear that schools have autonomy to decide:

- how and when subject matter is delivered
- how, when, and why learning experiences are developed, and the context in which learning occurs
- how opportunities are provided during study for explicit and integrated teaching and learning of complementary skills

Evidence about the quality of QCAA syllabuses was collected through an analysis of QCAA documents and judged against the criteria proposed by van den Akker et al. (2009) and shown in Table 4.3. The QCAA syllabuses include a detailed rationale, explicit aims and objectives, and content that includes descriptions of what students are expected to be able to do with the content. They provide extensive information about assessment and make clear its alignment with the aims and objectives. Time allocations are also included, as well as general guidance on learning activities, balancing guidance with freedom for teachers to adapt to the needs and interests of their students. Guidance is also provided about the role of the teacher, again balancing direction and freedom. Those syllabuses where fieldwork is required also address the selection of location and many address aspects of materials and resources available through digital technology and/or texts, including information about materials allowable for use in assessments.

Guidance is also provided about the role of the teacher, again balancing direction and freedom. Those syllabuses where fieldwork is required also address the selection of location and many address aspects of materials and resources available through digital technology and/or texts, including information about materials allowable for use in assessments.

The QCAA syllabuses well satisfy all 10 quality criteria proposed by van den Akker et al. (2009), as summarised in Table 4.5. Internationally, curricula and syllabuses which offer broad frameworks with much greater responsibility passed to schools and teachers commonly satisfy only three or four of the ten (van den Akker et al., 2009).

Quality criteria for syllabuses	Judgement of QCAA syllabuses against criteria
Rationale	Clearly provided
Aims and objectives	Clearly provided
Content	Clearly provided
Learning activities	Helpful suggestions provided
Teacher role	Discretion and professional responsibility clear
Materials and resources (learning activities)	Electronic links to electronic and print provided
Grouping	Suggestions made where relevant
Location	Suggestions re location provided where relevant
Time	Clear time limits proposed
Assessment	Specification of various assessments to be used

Table 4.5 QCAA syllabuses judged against international criteria

Evidence about syllabus development was collected through QCAA documents and focus groups and evaluated against these criteria. Most school staff data were collected during the early implementation stage of revised applied subject syllabuses in 2024, and before the introduction of new general subject syllabuses in 2025. Reactions to the 2025 general subject syllabuses were not available, although it is clear that the 2025 Essential English syllabus addresses concerns raised about its predecessor.

Planned and systematic development is evidenced in QCAA documents, which show a consistent process across subjects, requirements for literature reviews and legislative inputs, specified timelines and evaluation mechanisms, and the use of design briefs and templates. Consultations involved multiple stakeholders, critical friends, reference groups and committees.

Syllabus development was a core change-management element of the new QCE system, designed to build teacher confidence. Expert Writing Teams (EWTs) and Learning Area Reference Groups (LARGs) comprising practising teachers were recruited to support development and act as champions for change. Draft syllabuses were released incrementally so teachers could digest each element in stages (e.g., unit content, then assessment). This marks a substantial shift from the approaches of the Radford (1970) and ROSBA (1978) reports.

Another feature of well-structured curricula is the integration of components in coherent documents that set clear expectations for content, instructions and assessments. A review of all syllabuses in English (6), Mathematics (5) and Physics showed high consistency in structure and terminology. Glossaries and other supporting resources were also available to assist teachers and schools.

Survey responses indicated that teachers welcomed the higher level of definition, which increased consistency of teaching and learning across Queensland. QCAA staff working with schools and teachers using the syllabuses reported little negative reaction to the new resources and supports. Where curricula are based on senior secondary curricula in the Australian Curriculum, the QCAA syllabuses provide consistent detail with all other QCAA syllabuses.

Inclusive and consultative development was evident through multiple levels of consultation and the iterative nature of the processes. Development was led by curriculum professionals, particularly QCAA Principal Education Officers, who supported EWTs using design briefs and the Marzano and Kendall (2007) taxonomy. The taxonomy underpins the syllabuses and is visible in unit objectives and assessment information, providing alignment across curriculum, pedagogy, and assessment.

The cyclical nature of development is evident in the iterative design process, with multiple rounds of design and consultation taking place for each syllabus. While there is no fixed revision schedule, monitoring of implementation identifies when revisions are needed. State-wide consultation began in 2021 for applied subjects (implemented in 2024) and in 2022 for general subjects (implemented in 2025). This shows early commitment to cyclical review and revision.

Teachers expressed concerns about the rollout of the applied 2024 syllabuses, noting that resources had not been available early enough and comparing it unfavourably with that of the 2019 general syllabuses. Although the applied syllabuses were provided 12 months in advance, with supporting resources released progressively, some teachers wanted more time.

Stakeholders also raised concerns about teacher workload. The revision of the applied and general syllabuses coincided with the implementation of Version 9.0 of the Australian Curriculum (ACv9.0; ACARA, undated-a) which Queensland adopts as written. This compounding effect likely increased perceptions of workload, particularly for teachers involved with multiple subjects under revision.

Sustainability of the curriculum was another area of concern. Teachers cited the volume of content, timing of the implementation and availability of resources as concerns. It is easy to have unrealistic expectations of what can be achieved with available personnel, time and finances, so there will always be a balance to strike. The current cycle is reasonable, provided the changes build on what teachers have already learned to do with the current syllabuses.

Overcrowding of the curriculum and increased difficulty and rigour in the content were concerns. An over-crowded curriculum makes learning difficult for both teachers and students. For teachers, a crowded curriculum challenges the quality of their teaching, reduces flexibility to respond students' needs and can diminish their enjoyment of the profession. The introduction of a new curriculum also inevitably carries a sense of crowdedness because teachers lose familiar structure and content.

The changes to the curriculum were reported by teachers to have increased difficulty and rigour of the subjects. These comments were more common in mathematics and science, where many teachers, school leaders and parents suggested that the content was too complex and aligned more closely with university-level expectations. Since Queensland adopted the Australian Senior Secondary Curriculum in mathematics and science, the increased difficulty and rigour may be a consequence of that change.

Overall, against five criteria for high-quality curriculum development and revision shown in Table 4.5, the evidence shows that QCAA has a good approach to the syllabus development. The conclusions are summarised in Table 4.5. Some participants reported positive reactions, particularly in relation to increased clarity, improved assessment and better alignment with industry expectations. Teachers appreciated the detailed information provided in syllabuses, as it offers guidance about teaching and learning that aligns with international best practice and meets classroom needs. This has helped improve the consistency of teaching and learning across Queensland, as required by the Queensland Government (undated).

The reintroduction of External Assessments has imposed deliberate restrictions on how syllabus can be interpreted. While many saw this as contributing to comparability, some teachers regretted the increased prescription, preferring earlier arrangements that allowed more diversity of interpretation and implementation, albeit with less comparability across schools.

Requirement	Description	Queensland curriculum
Planned and systematic	Ensures that key content and curriculum priorities are mapped across subjects and domains. There is a strategic approach to the development and presentation of syllabuses resulting in a comprehensive curriculum	QCAA syllabuses satisfy this requirement
Inclusive and consultative	Showcases a diversity of perspectives with consultations from key stakeholder groups, particularly teachers. Including teachers in curriculum processes will allow teachers to feed their expertise into the process, as well as develop their own competencies and practices	QCAA syllabuses satisfy this requirement
Led by curriculum professionals	High-quality curriculum development is led by professionals who are experts in content areas as well as approaches to teaching and learning	QCAA syllabuses satisfy this requirement
Cyclical in nature	It is also cyclical in nature by incorporating mechanisms for ongoing review and revision	QCAA syllabuses satisfy this requirement
Sustainable	High-quality curriculum development is planned in ways that are sustainable for implementation in classrooms and schools. Thus, curriculum developers need to be mindful of the feasibility of teaching the planned curriculum within the context of development	Need to monitor any ongoing concern among teachers about timeline and workload involved in syllabus revisions

Table 4.6 Findings on QCAA syllabuses satisfaction of requirements for quality syllabus development

4.3 The special case of 21st-century skills

4.3.1 International focus on 21st-century skills

In the last few decades, there has been increased focus on developing what have been called students' 21st-century skills. These skills aim to serve students well into their futures, extending beyond the basic acquisition of knowledge. They are intended to prepare students for an uncertain future, where they will be able to apply their knowledge to solve problems they have not yet experienced. Skills such as how to learn, becoming life-long learners and applying knowledge in new contexts have been the goal of nearly 30 years of work led by the OECD (OECD, 1999, 2019).

In the Australian Curriculum, the term 'general capabilities' is preferred to 21st century skills for two reasons. First, many of the skills are not unique to the 21st century, with digital literacy perhaps being the only exception. Second, many of the skills are domain-specific rather than transferable and generic. For example, critical thinking and problem solving are general terms that can be used to name important skills, they have specific meanings in different domains of knowledge. For example, critical thinking and problem solving in science are not the same as critical thinking and problem solving in history. For this reason, the Australian Curriculum emphasises that the skills should be developed within subject domains. In the latest version of the foundation document The Shape of the Australian Curriculum Version 5 (ACARA, 2020) the place of general capabilities is explained in the following terms:

> *Discipline-based content knowledge is important for the development of general capabilities. General capabilities are developed drawing on the content of the learning area curriculum, rather than as isolated generic skills. The teaching of learning area content will be strengthened by the application of relevant general capabilities, as will the development of the general capabilities through appropriate learning area contexts.* (p. 16)

4.3.2 QCE treatment of 21st-century skills

The Shergold Report (2020) observes that:

> *The Australian Curriculum currently includes the teaching of a range of important general capabilities from Foundation to Year 10. Yet, ironically, this focus on attributes that are seen as important by employers is not made explicit within the senior secondary years. Rather, the greater focus is on scholastic achievement in the form of subject scores and, most particularly, the Australian Tertiary Admission Rank (ATAR).* (p. 35)

While that may be generally true of senior secondary curricula in other States and Territories, it is not the case for the Australian Curriculum senior secondary curriculum subjects implemented in Queensland. Nor is it true of QCAA syllabuses, all of which pay attention to 21st-century skills embedded in specific subjects in recognition of their domain-specific nature and limited generalisability.

Despite the different terminology, the general capabilities in the Australian Curriculum are consistent with what QCAA syllabuses and elsewhere refer to as 21st century skills. A QCAA position paper (2024a) explains how its integration of the 21st century skills build on those embedded and developed through the Australian Senior Secondary Curriculum general capabilities:

> *The Queensland Curriculum and Assessment Authority (QCAA) has identified and defined a set of 21st century skills based on national and international research about the skills students' need in the 21st century. Along with literacy and numeracy, these 21st century skills are the underpinning factors shaping the development of the senior syllabuses. These skills build on those that are embedded and developed through the Australian Curriculum general capabilities in prep to Year 10 (QCAA, 2022a). These 21st century skills will help prepare Queensland students by providing them with the knowledge, skills, and confidence they need to be equipped for the demands of higher education, work, and life, and to participate effectively in the community and the economy in a complex and rapidly changing world (QCAA, 2024a, p. 1).*

Findings

- There is strong general approval of the quality of the new syllabuses and, with applied subjects, of the alignment with industry needs. They have a sound base in research findings
- The one clear exception to this general approval was 2019 Essential English, which did not align well with the other English subjects. It was constrained by two requirements: building on students' Year 10 achievement levels developed with the Australian Curriculum, and seeking to ensure students reach at least Level 3 of the Australian Core Skills Framework (ACSF; McLean et al., 2012)
- QCE curricula satisfy the quality criteria recommended by OECD. QCAA's syllabus development involves systematic, wide communication with subject and teaching experts and the broader community, reflecting best practice in incorporating advances in knowledge

Recommendations

- That QCAA prioritise a review of the implementation of the Essential English (2025) curriculum and syllabus.
- QCAA to establish a research agenda to identify factors that might require amendment to current syllabuses and then monitor the implementation of changes, including the impact on timeline and workload

4.4 Supporting policies, procedures and systems

Area for review 5.1

Whether the policies, procedures and supporting resources that comprise the QCE system are fit-for-purpose. This should include an assessment of whether the qualification is contemporary and supports students to follow their chosen pathways.

4.4.1 Open availability of information

QCAA's website (www.qcaa.qld.edu.au) provides full information on the QCE and the QCIA including all information contained in the QCE and QCIA Policy and Procedures Handbook Version 6.0 (QCE & QCIA Handbook, 2025). The site is comprehensive, easy to navigate and searchable. Except for information available only to those with accounts to access materials through a restricted Portal (typically teachers, school leaders, and assessors), all material is publicly available.

4.4.2 Flexible pathways for students

Although support systems existed before the introduction of the new *QCE* in 2019, several key features for students are worth noting.

1) Flexible pathways: Students can combine general subjects, applied subjects, recognised studies, university subjects and VET qualifications. The new QCE system also recognises prior learning in two ways: (a) learning completed before moving to Queensland, and (b) learning recognised by a Registered Training Organisation (RTO), either school-based or external (QCE & QCIA Handbook, 2025, pp. 183–184). The *myQCE* website helps students plan their subject selections to achieve a (QCAA, 2025e).

- Individual learning accounts: These provide students with a digital record that tracks their progress toward QCE attainment and allows schools to track credit accumulation, identify students at risk of non-completion and provide support (QCAA, 2023e, 2023j).
- Literacy and numeracy: Literacy and numeracy skills remain requirements for QCE eligibility. These are embedded in redeveloped subject syllabuses, and short courses are also available with more targeted support (QCAA, 2018; QCE & QCIA Handbook, 2025).

The new QCE system supports students to pursue their chosen pathways. The breadth of learning opportunities makes the qualification flexible, allowing students to follow their interests and prepare for further study and work. Students can take both an applied subject and a VET qualification, but they cannot duplicate credit for overlapping courses. For example, students may study Hospitality Practices and also complete a Certificate II in Hospitality, but they can accrue only four credits towards their QCE across these studies. Both studies will still appear on their Senior Statements, the official record of achievement for Year 12 students (QCE & QCIA Handbook, 2025).

Students' planning for their senior secondary years commences in Year 10 with students and parents developing Senior Education and Training (SET) Plans (QCAA, 2023h). This practice commenced prior to the introduction of the QCE in 2008. Survey responses for this evaluation indicated improved alignment between SET Plans and the new QCE requirements.

Offering this breadth of options with a single qualification is a strength, and unusual in international terms. In many countries (e.g., Poland, Norway, Germany), different qualifications are awarded to students depending on whether students take primarily academic or vocational subjects (Teo et al. 2023). In Australia, breadth has sometimes been achieved through separate qualifications. For example, Victoria once had four different senior certificates (McGaw & Hannan, 1985, pp. 66–68). These were replaced by a single Victorian Certificate of Education (VCE) in 1985, followed later by the Victorian Certificate of Applied Learning (VCAL). VCAL ran from 2002 to 2023, when it was replaced by the VCE Vocational Major and the Victorian Pathways Certificate (Victorian Government, 2025).

Diversity of offerings at a state-wide level, however, does not always translate into diversity at the school level. For example, when the NSW senior curriculum was reviewed in the 1990s, the standard pattern was for a subject to include a 3-unit course, a 2-unit advanced course (consisting of two of the units from the 3-unit course), and a 2-unit general course. In English, there were four versions, with a 2-unit Contemporary English course positioned below the 2-unit general course.

At the time, NSW had state-wide assessment of students in English in the School Certificate at Year 10 and the Department of Education maintained a strong database that tracked students into upper secondary education. Comparing students in non-selective government high schools in the more advantaged metropolitan north and the less advantaged metropolitan southwest regions revealed (as shown in Table 4.7) that among students in the top 10% in School Certificate English, those in metropolitan north were more likely than those in metropolitan southwest to enrol in either of the two more demanding English course in Years 11–12. A similar pattern was seen for students in the top 30% at Year 10 (McGaw, 1997, p. 44).

	Top 30% in 1994 School Certificate English				Top 10% in 1994 School Certificate English			
1996 HSC English Course	Metropolitan Southwest		Metropolitan North		Metropolitan Southwest		Metropolitan North	
	M	F	M	F	M	F	M	F
English 3 Unit & 2 Unit Advanced	13	22	22	35	33	53	56	63
English 2 Unit General	78	70	73	62	67	48	44	36
Contemporary English	9	8	5	4	0	1	0	1

Table 4.7 Percentage of NSW government non-selective school students in different English courses

The reason for the discrepancy in Year 11–12 enrolments between students with similar Year 10 achievements was that schools did not provide all options in Year 11–12. In metropolitan north, 59% of the schools offered 3-unit English while in metropolitan southwest only 38% did. Taking the 3-unit English and 2-unit Advanced English together, the percentage of schools offering both was 89% in the metropolitan north and 62% in the metropolitan southwest.

The QCAA website provides information on the subjects offered by each school together with the number of students enrolled in each unit. The display is ordered by subject so users can see which schools offer each subject. The data can also be arranged to show the subjects offered by individual schools. However, the site does not provide information on how restrictions in subject offerings within schools may affect student choice, of the type revealed for NSW in Table 4.1. This is an area that QCAA could consider monitoring – for example, by using students' NAPLAN results to examine whether students with similar prior achievement levels have similar study options for the QCE regardless of the school they attend. Excessive diversity at the state level that cannot be replicated within individual schools risks reinforcing disadvantage, as it did in NSW in the 1990s.

4.4.3 Supporting policies made accessible for a variety of stakeholders

The Senior Education and Training (SET) planning process is highly regarded by students and parents. Parents and parent associations valued working with teachers throughout the planning process, noting that school staff understood the learning options available in their context as well as their student's capabilities and interests. Students generally described SET planning as a positive experience, although some reflected that their plans changed as they progressed through Years 11. This flexibility was seen as strength of the system.

School staff reported that they often drew on QCAA resources to support SET planning and valued these as a starting point for tailoring information to parents and students. A common difficulty raised was determining the credits associated with some students' plans, particularly where there was confusion about duplication between school-based apprenticeships and traineeships. While participants were mostly positive about the duplication policy that prevents students earning multiple credits for the same learning outcomes in different courses, they suggested greater clarity on the number of credits available for apprenticeships and traineeships would improve the process.

In school case studies, stakeholders also noted that the QCE credit allocation did not always reflect the amount and difficulty of student work. The most common concern was that the credit awarded for VET Certificate II was excessive compared with general subjects, with some also questioning the allocation of eight credits for a VET Diploma. QCAA's credit allocation for VET studies is based on nominal hours set by the National Centre for Vocational Education Research (NCVER). These policies pre-date the 2019 reform, but alignment between credit allocation and the actual workload remains important. Western Australia has recently moved in this direction, announcing that Diploma-level VET qualifications will no longer receive unit equivalence towards WACE achievement, a change said to "support quality outcomes for students undertaking VET as part of their WACE studies" (School Curriculum and Standards Authority, 2025, p. 6).

Overall, the flexible pathways and SET planning process are seen as positive components of the new *QCE* system, helping students prepare for and pursue a variety of post-school pathways. The accessibility of supporting materials also demonstrates the transparency needed to build confidence in the system.

Findings

- The wide subject offerings of the new QCE system allow students the possibility of pursuing their chosen pathways with the SET planning process effectively, supporting discussions with teachers and decisions by students and their families
- The duplication policy preventing students from earning multiple credits for the same learning outcomes in different courses is well-regarded

Recommendations

- QCAA to review how much of the diversity of QCE offerings at the state level can be provided to students within schools and determine whether any limitations in capacity at the school level disadvantage particular groups of students
- QCAA to review the credit awarded for VET Certificate and Diploma courses in comparison to subjects with syllabuses developed by QCAA

05

Assessment processes

5. Assessment processes

5.1 New assessment processes

Area for review 4.1

Whether the key elements of the new assessment system are fit-for-purpose, and their theoretical basis strengthened through implementation. The focus should be on the three main assessment processes of Endorsement, Confirmation and External Assessment.

Factors to consider are the interrelationships between the three processes, educational benefits to students, success in improving validity and reliability of assessments, value for money and return on effort invested in their delivery.[1]

This chapter presents evaluation findings regarding the fourth and fifth elements of the new QCE system identified in the terms of reference.

Broadly, the reform of senior secondary education in Queensland aimed to improve the balance between tensions that exist in all secondary education systems. These tensions lie between the need for comparability of results across the system and the importance of allowing appropriate local contextualisation of instruction and assessment. Teachers, as those closest to the students and their communities, are best placed to understand individual learning progress and identify the next steps within each student's zone of proximal development (Vygotsky, 1978).

In any assessment system, several goals must be met: incorporating local context; generating reliable evidence of student capability and achievement; ensuring students have a solid grounding in established knowledge bases; and drawing on multiple sources of evidence to support valid inferences about the knowledge, skills and abilities of the holders. No single assessment can or should support all of these. Rather, it is the combination of coherent instruction with high-quality formative and summative assessments that provides the necessary evidence base to support the desired inferences.

QCAA has designed the QCE assessment system to address these goals through three monitored components: Endorsement, Confirmation and External Assessment.

- **Endorsement** gives QCAA assurance that the Internal Assessments contributing to student results have high contextual validity. Designed by local teachers, these assessments also utilise the teacher's knowledge of their students to maximise accessibility
- **Confirmation** ensures that the marking of the endorsed Internal Assessments aligns with the instrument-specific marking guide, maximising reliability
- **External Assessments** are developed and marked by QCAA to provide a common, external assessment for the relevant general subjects. The assessments are administered under standardised conditions for each subject, including on the same day and at the same time across the state, to enhance the reliability, validity and comparability of the student results.

[1] QCAA and UoM agreed value for money and return on investment could not be substantially addressed in this engagement. QCAA should consider an evaluation of these elements as part of its long-term program monitoring.

Each component contributes to the final student result and award of the QCE, and each provides strong evidence for one or more of the QCAA-defined attributes of quality assessment. Taken together, Endorsement, Confirmation and External Assessment create a profile of evidence with high levels of validity, reliability and accessibility. While the components may not be directly interrelated, each has a primary focus, and together they support valid inferences about each student's knowledge, skills and abilities.

QCAA completes quality assurance of the QCE certificates. It describes assessment as:

> Assessment plays an integral role in improving learning and informing teaching. Its fundamental purpose is to establish where learners are in an aspect of their learning at the time of assessment. (QCAA, 2025h)

Key considerations include whether assessments provide challenging opportunities for students to demonstrate what they know and can do; whether they align accurately with the syllabus and learning objectives; whether they are reliable across contexts; whether results are comparable; and whether the overall evidence supports valid inferences about student knowledge, capability and progression.

The ACER review by Matters and Masters (2014) thoroughly assessed the system at the time and made a series of recommendations for significant change, grounded in international research on educational systems and assessment design structures. The recommendations aimed to:

- streamline and simplify the senior secondary education and awarding system in Queensland
- increase transparency and public understanding
- improve comparability and reliability of the assessment results
- strengthen the validity of the inferences based on those results

One of the most debated areas in the ACER review (Matters & Masters, 2014) was moderation of results and the role of External Assessment. The challenge was how to include an external examination without privileging its role over the Internal Assessments. Misunderstandings often arose from the assumption that incorporating externally set exams meant statistical moderation. Supporters of statistical moderation saw it as a way of 'fixing' incorrect teacher judgements, while opponents viewed it as a mark of mistrust that favoured external examination over teacher-set assessments. The balance reached included the collection of evidence from both teacher-developed assessments and externally set and marked assessments in determining the subject result. For the Internal Assessments, consistency and quality across the state would be supported through standardised curricula, common marking guides and Endorsement and Confirmation by the authority. In this way, local contextualisation by teachers and QCAA's quality assurance both contribute to the final student outcome. With External Assessment contributing either 25% or 50% of the student result, the system does not privilege external exams over teacher assessments.

Moderation approaches in other Australian states

Other states in Australia use varied approaches to moderate their Internal and External Assessment information, supporting comparability across schools within each state. National comparability for tertiary entrance purposes is presumed through use of the ATAR, although ATARs are constructed differently in each state. Moderation serves two purposes: to provide comparable and accurate reporting of senior secondary results, and to support calculation of the ATAR. In terms of moderation processes within the states:

- The Australian Capital Territory (ACT) uses qualitative moderation to produce senior secondary student results. There are no subject-based External Assessments used for this process. The moderation uses peer review of portfolios of student work against system-wide criteria that are common to all schools. ACT uses a scaling test (AST) to align student results, combining the results from the student's best four course scores and the AST results to support calculation of an ATAR.
- New South Wales (NSW) issues a qualification comprising internal school-administered and external state-wide standardised exams, each worth 50% of the final mark. Statistical moderation in NSW adjusts school assessment marks based on the HSC exam, which is given under common conditions.
- South Australia (SA) and the Northern Territory (NT) use a common procedure and certificate. Their process uses a stratified sampling approach and the judgement of expert moderators to evaluate the student work and teacher-assigned grade, with a proportion of grades adjusted to align with the standards.
- Tasmania uses statistical moderation through external exams set and graded by the Tasmanian Assessment, Standards and Certification (TASC) authority with statistical procedures for anomaly detection. Social moderation is used for selected criteria and tasks.
- Victoria uses statistical moderation through external exams set by the Victorian Curriculum and Assessment Authority (VCAA) and, in some cases, the General Achievement Test (GAT).
- Western Australia (WA) awards a certificate of education based on school marks. Comparability is established through basing instruction and assessment on a common syllabus, use of the authority's grading procedures, and participation in the consensus moderation activities, including syllabus delivery audit, consensus moderation and reviews of documentation and grading. The results are statistically scaled against the External Assessment. In WA, each ATAR subject score is determined by combining 50% from the student's school results and 50% from the Western Australian Certificate of Education (WACE) examination.

Overall, most Australian students receive results that are statistically moderated to adjust school assessments for tertiary admission.

The Queensland senior secondary system is grounded in the literature on formative and summative assessment. Based upon contemporary criteria for world-class assessment systems (OECD, 2024), the new system incorporates flexibility for students, transparency in assessment processes and different opportunities for students to demonstrate their knowledge, including choice and multiple pathways to qualifications. It balances school-based formative assessments, administered during regular instruction, with summative assessments defined in policy and syllabuses. It is recognised, however, that what students, teachers, school leaders and other stakeholders experience is not always fully aligned with the written curriculum and policies.

Current thinking on world-class assessment systems reflects more than 40 years of international research into how students learn. This section outlines best practice in assessment design, including the role of formative and summative assessment, transparency and support for assessment procedures, and assessment methods. Learning culture, although clearly important, is outside the scope of this evaluation. Supporting evidence for the alignment of the QCE system with these principles is presented later in this chapter, along with recommendations in areas where practice could be strengthened.

An aspect of particular interest is that the operationalisation of the system strongly parallels a best-practice concept in Black and Wiliam (2018): the key distinction between formative and summative assessments lies in the inferences drawn from the assessments, not the assessments themselves. Any assessment can be used formatively, summatively or both. Evidence gathered in our review of the new QCE process suggests that it is a contemporary qualification that is aligned with components of a high-quality assessment system. A well-designed system contains a balance of formative and summative assessments that are organised and scaffolded appropriately; such a balance is embedded in the operational QCE system that exists today.

The roots of the ideas in Black and Wiliam (2018) go back decades. In Tyler's (1949, p. 1) classic book, he proposes four fundamental questions to be answered in developing curricula and instructional plans:

(1) *What educational purposes should the school seek to attain*
(2) *What educational experiences can be provided that are likely to attain these purposes?*
(3) *How can these educational experiences be effectively organised?*
(4) *How can we determine whether these purposes are being attained?*

Black (2016, p. 728) offered a similar model, with Tyler's (1949) question 4 expanded into two steps, separating the evaluation of attainment into review and formal assessment:

Step 1 – Formulating clear aims
Step 2 – Planning activities
Step 3 – Implementation and Interaction in dialogue
Step 4 – Review of the learning achieved
Step 5 – Formal summative assessment

In both models, assessment is understood as an integral part of curriculum and pedagogy – indeed, as part of learning itself. Many frameworks in this space focus on the first three steps (often described as 'instructional practices') but do not delve into the evaluation of effectiveness through integrated assessment practices. Black and Wiliam (2018) extend this view by presenting the process as cyclical rather than sequential, emphasising interaction, feedback and backtracking based on evidence gathered at each stage. Their model is shown in Figure 5.1.

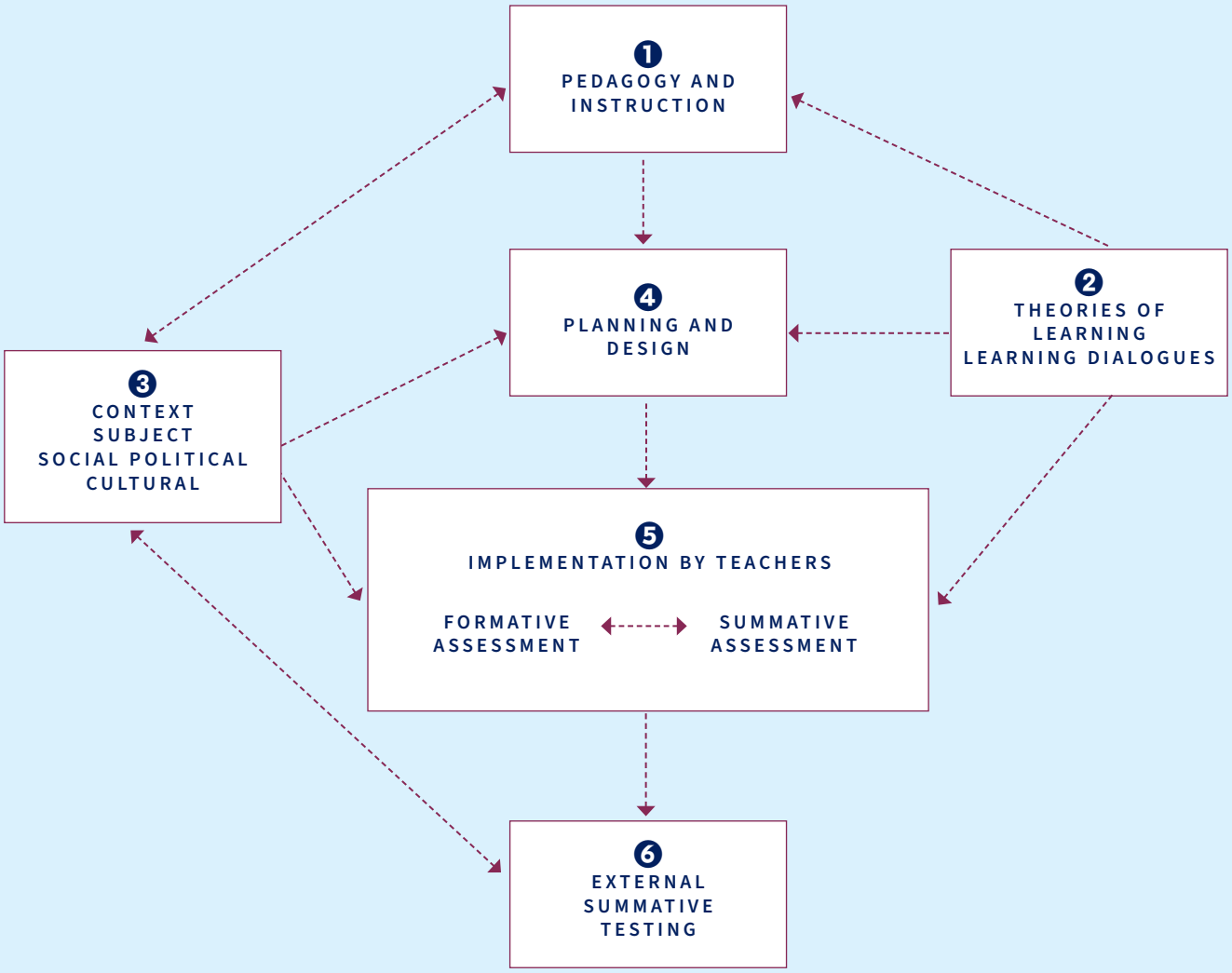


Figure 5.1 Model for assessment in relation to pedagogy

(Adapted from Black and Wiliam (2018))

Specifically, curriculum development in Queensland is a collaborative process that draws on the expertise of subject-matter experts and teachers. The cycling between steps 1, 2, 3 and 4 is embedded in the establishment, review and renewal of the curricula. Queensland's assessment processes also align closely with steps 5 and 6. Teachers develop assessments as part of their regular classroom instruction, and these can be used formatively, summatively or both. When assessments are intended for summative use, those in Units 3–4 must align with the established curriculum and the ISMG. These assessments undergo formal Endorsement by QCAA, and the application of the ISMG in marking is reviewed in Confirmation. Teacher-developed assessments used solely for formative purposes are not examined by QCAA in these formal stages, but they too will be guided by the curriculum. This allows students to assess their own learning and that of their peers, engaging in purposeful and interactive dialogue. The teacher serves as a guide for these discussions, ensuring they remain focused and aligned with the intended learning outcomes. The QCAA-approved curriculum and ISMG provide concrete developmental guidance that helps align students' understanding of 'quality work' with their teacher's expectations (Sadler, 1989) and QCAA standards.

In reforming the senior secondary QCE system, Queensland sought to balance competing demands: maintaining centralised control while preserving instructional autonomy. The current QCE system achieves this balance, as illustrated in Table 5.1 for some specific aspects of the system.

Table 5.1 Control and autonomy in QCE system components

System Aspect	Control	Autonomy
Instructional delivery	Content and learning objectives defined in authority-provided syllabus	Teachers and schools have latitude to develop lesson plans and instructional practices appropriate to their local context
Development and marking: formative assessment	Assessment literacy courses offered to all senior secondary teachers designed to improve all assessments administered in Queensland schools	Teachers and schools have latitude to develop in-stream formative assessments appropriate to their local context
	Feedback provided to teachers and schools from the Endorsement and Confirmation processes, as well as annual subject reports with examples of effective practices and areas requiring strengthening	Feedback from the Endorsement and Confirmation processes helps build teacher capacity to design, deliver and make judgements about high-quality assessments
Development: summative Internal Assessment (Endorsement)	Endorsement procedures assure that summative assessment instruments align with syllabus objectives, specifications and conditions	Teachers collaborate within their school to develop assessment instruments that align with the syllabus but retain the flexibility to contextualise to local settings
Marking: summative Internal Assessments (Confirmation)	Confirmation procedures ensure consistency in application of the ISMG across cohorts and schools and comparability of subject results across the state	Student responses are marked locally, with a stratified sample provided to QCAA for Confirmation. If the sample marks are confirmed, all other student marks are also confirmed
External Assessment	External Assessments are developed, administered under standardised conditions and marked by QCAA	Teachers prepare students for the external exams using their professional judgement as to the most effective approaches in their local context. External Assessments are developed by a team of trained subject matter experts, including teachers and academics. Teachers are also encouraged to apply for marking and other roles needed in the assessment process

Another tension is the balance between consistent assessment of student learning (referred to as 'comparability' here) and personalisation. Comparability has been a touchstone of the field in formal, and especially standardised, assessment for more than 100 years. Comparability is associated with complex and challenging issues, although over time many of these issues have been addressed and have well-understood methodologies that are applied in large-scale assessment systems (Berman et al., 2020). More recently, there has been a move towards personalisation of assessments (Bennett, 2024), as technological advances make that goal increasingly feasible. The current QCE system is designed and delivered with a strong focus on comparability, while still allowing local context and autonomy. In this balance, the system is well placed, with scope for further progress.

While assessments are not yet personalised at the level of each student, they are tailored at classroom and school levels. In some areas – such as Physical Education, Arts or Design – assessment is already more finely personalised. The Endorsement process allows flexibility in design and content within boundaries of the approved curriculum and marking guides. Teachers can set formative and summative assessments within the appropriate social, political and cultural contexts while maintaining subject and discipline standards. Confirmation brings comparability back into the frame, ensuring that the approved assessments are marked against a common guide in a consistent manner. The inclusion of an External Assessment provides further assurance that student results are both individualised and comparable.

Another notable aspect of the QCE system is its strong focus on assessment literacy and support for teachers' assessment practices. Research (Maclellan, 2004; Volante & Fazio, 2007) has shown that beginning teachers lack the skills and knowledge required to measure and communicate student learning effectively. This gap persists despite consistent evidence that practices such as questioning, diagnostic assessment, peer feedback and group assignments improve student learning outcomes (DeLuca, 2012). While formal coursework in assessment improves teachers' ability to describe various modes and purposes of assessments, it does not always prepare them to address practical issues of reliability, comparability, validity and equity (Maclellan, 2004).

QCAA addresses this gap by providing a comprehensive sequence of assessment literacy modules in the QCAA portal. Teachers applying for assessor roles must complete a specified subset of these modules successfully. More importantly, the assessment design structure requires teachers to apply what they know about assessment to building instruments aligned with a syllabus, and provides feedback and support if they fail to demonstrate the attributes of quality assessment. Accuracy in applying marking guides is evaluated in Confirmation, with similar feedback and support structures in place. For those interested in advanced applications, there are course opportunities for developing and marking the External Assessments. This combination of formal training and practical application supports a teaching workforce that could be expected to have unusually strong assessment literacy and practices.

A system that supports both control and autonomy, personalisation and comparability, and that equips teachers and schools to deliver timely, reliable results with valid evidence of student capability represents a successful design.

As part of this evaluation, the Endorsement, Confirmation and External Assessment processes were explored via a range of data sources:

- QCAA document and website review
- assessment data sets
- psychometric analysis of item and assessment functioning
- evaluator observations
- review of prior evaluations and other internal and external reports
- surveys both targeted to specific stakeholder groups and more generally
- written submissions and survey comments
- discussions with key stakeholders in focus group sessions
- individual interviews
- case studies at selected schools

Key findings for each assessment are presented below, including major themes that emerged across evidence sources.

The purposes of the psychometric analyses, including the statistical or psychometric models, analytic approaches, and results for each sampled subject are provided in Appendix B**.**

5.2 Endorsement

Endorsement is one of the annual quality assurance processes applied to summative Internal Assessment instruments in Essential English, Essential Mathematics and all general subjects, including general extension subjects. Its purpose is implicitly framed as building confidence among schools and students in the quality of Internal Assessments (IAs), while also refining teachers' assessment design skills. The QCAA website does not clearly articulate this intent, although it notes that Endorsement evaluates the extent to which assessments demonstrate two attributes of quality assessment: validity and accessibility. Both terms are defined on the QCAA website and in the Endorser Manual (QCAA, 2024i), although the definitions vary somewhat. In practice, Endorsement adds to the body of evidence supporting valid inferences about student competence in a subject area, with a particular focus on construct validity. Accessibility, as defined by QCAA, aligns with fairness in the Standards for Educational and Psychological Testing (American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME), (AERA et al., 2014)), although the broader concept of fairness, with its emphasis on removal of barriers to performance, is less clearly addressed in the Endorsement process.

While validity and accessibility are identified by QCAA as key characteristics of Endorsement, the process also contributes evidence of reliability, particularly in ensuring comparability across multiple forms of an assessment. The syllabus and ISMG are set in advance for all Internal Assessments submitted for Endorsement in Units 3–4. Endorsement verifies that the assessments submitted by schools across the state function as 'alternative test forms' designed to assess the same construct, and that each is evaluated against the same marking guide standards.

When assessments are not endorsed, one or more priorities under validity, accessibility or both is identified as the reason in the subject report for the cohort in that year. The validity priorities are:

1. alignment
2. authentication
3. authenticity
4. item construction
5. scope and scale

The accessibility priorities are:

1. bias avoidance
2. language
3. layout
4. transparency

The priorities are defined and expanded in the assessment literacy courses that Queensland teachers are encouraged, but not required, to complete. These priorities are not outlined on the public QCAA website or in the Endorser Manual, although the manual notes that the attributes of quality assessment are further elaborated in the Endorser app. Access to these courses and materials, including detailed definitions, requires a valid login to the secure QCAA portal.

In 2023–24, approximately 27,500 school-developed assessments were reviewed by a casual workforce of more than 1000 QCAA assessors, most of whom were practising teachers (QCAA, 2024c) (see Figure 5.2). For the 2024 exit cohort, assessors endorsed 67% of assessments at first application in August 2023 and 65% in February 2024 (QCAA, 2024c), although Endorsement rate varied by subject and assessment technique.

A review of QCAA annual reports since 2020 indicates that first-round Endorsement rates have generally improved over time, suggesting that teachers are gaining competence in creating assessment instruments that meet the attributes of quality assessment.

That said, the 2024 rate still shows that only two-thirds of submissions were endorsed on the first attempt (QCAA, 2025c). Queensland's endorsement process is unique nationally, and there are no clear benchmarks for what constitutes an 'ideal' rate. Teachers have had five years of experience with the process and it would be desired that this would produce a higher initial rate than has been observed. The rate of Endorsement rises to 95% after resubmission, and all assessment instruments used with students have been endorsed prior to such use.

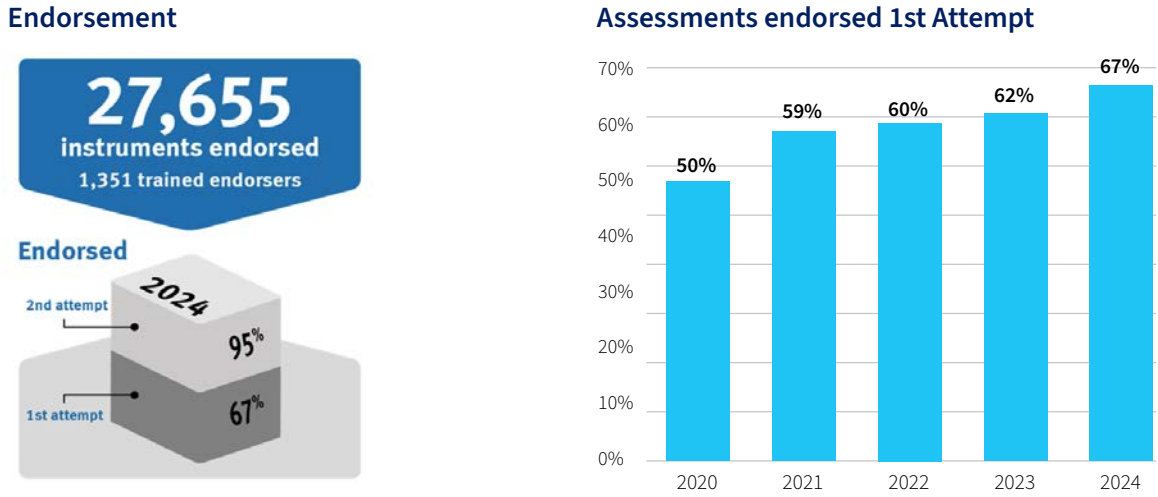


Figure 5.2 Assessment instruments endorsed at first attempt, 2020–2024

The Endorsement process is supported by a custom-designed Endorsement software application that contains a clear workflow and internal messaging functions to facilitate communication between endorsers, the Lead Endorser for their team and the Chief Endorser. The Endorsement app allows endorsers to contact the Lead Endorser directly with questions, while Lead Endorsers can communicate with team members about calibration, decisions or directives arising from the review. The system also enables escalation of instruments with unresolved issues to the Lead or Chief Endorser. Further detail, including a process map, is provided in the QCE and QCIA Handbook (2025).

The Endorsement app provides clear and well-designed support for a complex process and was noted by QCAA staff as a major factor in the successful transition to remote work during the COVID-19 pandemic. Staff also highlighted the value of its continuous improvement cycle, with new tools being introduced quickly and effectively. The app enables rapid communication between assessors, supports independent review of assessments and ensures smooth workflow and strong quality control of the Endorsement process.

The app workflow itself is straightforward. An endorser reviews a submitted assessment and determines whether it demonstrates the attributes of quality assessment (validity and accessibility). Instruments may not be endorsed if they use or closely replicate QCAA sample assessments or for other reasons based on the validity and accessibility priorities. If not endorsed, the instrument is returned to the school with identified priorities and written directives for improvement. The Lead Endorser reviews all Endorsement and non-Endorsement decisions and makes the final call, which is communicated to the school through the app. The process is repeated at application 2. If the instrument is still not successfully endorsed, Chief Endorsers and QCAA staff intervene directly, working with the school until the instrument meets the required standards of validity and accessibility.

Following each Endorsement event, the Chief Endorser contributes to a subject report. This report provides a comprehensive analysis of the Endorsement cycle for each subject, including data breakdowns, examples of effective practice and areas for improvement based on observations from the most recent cycle. These reports are made available to school leaders, subject leaders and teachers, supporting ongoing refinement of assessment practices in Queensland.

Qualitative evidence from endorsers was collected through surveys (843 surveys sent, 322 responses) and free-response comments (175 comments). Overall, endorsers felt that the current process is much better than the former QCE system and has improved since its introduction. Concerns raised included endorser workload, scheduling and potential impact of fatigue on judgements. Some endorsers requested more exemplar materials and commentary to guide their Endorsement and directive writing, while others recommended making pre-endorsed assessments available for school use. Both suggestions were seen as ways to reduce workload and increase work quality. More detail on the survey responses is available in Appendix E.

Findings

- Endorsement of the IAs assures assessment of the construct as defined in the syllabus and that student responses can be evaluated against the same marking guide standards
- All assessments are endorsed by QCAA prior to being administered to students in classrooms
- Endorsement rates have been increasing steadily over time although rates at first submission remain relatively low
- Annual subject reports contain data regarding the validity and accessibility priorities that led to non-Endorsement of assessments
- Each syllabus provides detailed information and guidance on the assessment objectives, specifications, conditions, mark allocation, and ISMG for each IA that supports the development of instruments that will fit the attributes of quality assessment
- Endorsed IAs can be expected to contribute solid evidence to support the validity of inferences about student achievement

Recommendations

- QCAA to analyse Endorsement rates annually within and across subjects for each validity and accessibility priority and examine the results for informative patterns for variables of interest such as geographic location and content area

Suggested operational adjustments

- QCAA to add clear definition of the intended purpose of Endorsement as well as list and define the validity and accessibility priorities on the public website
- QCAA to consider collecting additional data such as teacher experience level to enhance analysis of Endorsement rates
- QCAA to publish and explain the Endorsement rates for each cycle
- QCAA to review the Endorsement schedule and evaluate the feasibility, schedule impact and financial cost of a longer review period as well as workload of Lead and Chief Endorsers to ensure that sufficient time is allowed for planned training, operational activity and quality assurance tasks
- QCAA to design and implement a system for providing quality assurance data and feedback to all endorsers, to collect additional appropriate data to support such feedback and to ensure that Lead and Chief Assessors meet established minimum feedback expectations

5.3 Confirmation

Confirmation is one of the annual quality assurance processes applied to marking of summative Internal Assessment instruments for Units 3–4 in all general and general extension subjects. Confirmation targets the accuracy and consistency of application of the ISMG to student responses and supporting evidence provided for the summative assessment instruments by teachers across the state. The process increases the reliability of student results through increased comparability of marked results on the locally developed and endorsed assessments. This is an aspect of fairness as defined in the Standards (AERA et al., 2014) – specifically, fairness as part of test administration procedures. Although neither the assessments themselves nor the administration conditions are fully standardised, evidence supporting validity through content comparability was established previously during Endorsement. The use of common ISMGs enables Confirmation of consistent marking, contributing to evidence of reliability and fairness. Both are key aspects of the QCAA's attributes of quality assessment.

Confirmation is conducted on a stratified sample of marked student responses, selected to represent the full range of student performance. The sample is based on the distribution of achievement in the IA results and size of the cohort, with consideration also given to the school's Confirmation outcomes in previous events. Importantly, Confirmation is *not* a re-marking of the submitted work.

It is also important to note that ISMGs are constructed differently across subject areas, with teachers supported to understand these differences through a range of resources and assessment literacy courses. The ISMGs are designed to reflect requirements of different assessment techniques and conditions and therefore vary between subjects and Internal Assessments. Some features of the ISMGs are shown in the Table 5.2 below for a sample of courses across content areas. This variability in structure and expectations may influence the results of the psychometric analyses presented later in this report.

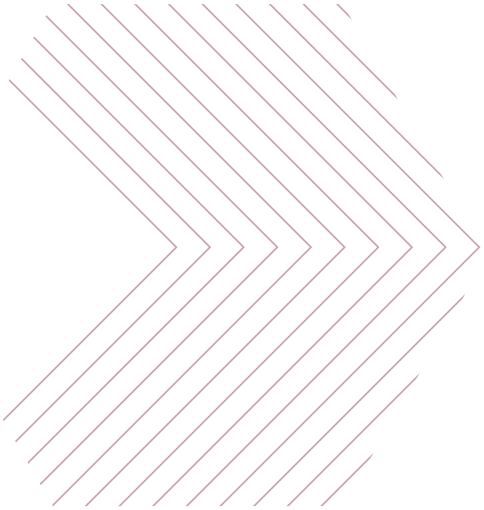


Table 5.2 Instrument-specific marking guides (ISMGs) rationale and design

Key Learning Areas (KLAs)	Sample subjects	Rationale	Instrument-specific marking guides (ISMGs)	Mark criterion and allocation
English	English English General 2025 v.1.2, (QCAA, 2024j)	The subject English focuses on the study of both literary texts and non-literary texts, developing students as independent, innovative, and creative learners and thinkers who appreciate the aesthetic use of language, **analyse** perspectives and evidence and challenge ideas and **interpretations** through the analysis and creation of varied texts.	• Internal Assessment 1: Spoken persuasive response (25%) ◦ focuses on the creation of a perspective through reasoned argument to persuade an audience. • Internal Assessment 2: Written response for a public audience (25%) ◦ responds to two texts connected by the representation of a concept, identity, time, or place, in a written response for a public audience. • Internal Assessment 3: Examination – extended response (25%) ◦ focuses on the interpretation of a text from the prescribed text list as a springboard for an imaginative response. • External Assessment: Examination – extended response (25%)	• Knowledge application ◦ creation of perspectives and representations of concepts, identities, times and/or places in the texts ◦ manipulation of the ways cultural assumptions, attitudes, values and/or beliefs underpin texts and invite audiences to take up positions ◦ use of aesthetic features and/or stylistic devices in the texts • Organisation and development ◦ use of the patterns and conventions of texts, and of the role of the speaker/designer, to achieve a particular purpose ◦ selection and synthesis of subject matter to support perspectives ◦ organisation and sequencing of subject matter to achieve a particular purpose, including discerning use of cohesive devices to emphasise ideas and connect parts of texts • Textual features ° language choices for purposes ◦ combination of a range of grammatically accurate/appropriate language structures to achieve purposes ◦ use of mode-appropriate features to achieve purposes
Mathematics	Mathematical Methods Mathematical Methods 2025 v.1.2, (QCAA, 2024l)	The major domains of mathematics in Mathematical Methods are Algebra, Functions, relations and their graphs, Calculus and Statistics. Topics are developed systematically, with increasing levels of sophistication, complexity and connection, and build on algebra, functions and their graphs, and probability from the P–10 Australian Curriculum. The ability to translate written, numerical, algebraic, symbolic and graphical information from one representation to another is a vital part of learning in Mathematical Methods.	• Internal Assessment 1: Problem-solving and modelling task (20%) ◦ provide a written response to a specific mathematical investigative scenario or context using subject matter from at least one of the topics in Unit 3 or Unit 4. • Internal Assessment 2: Examination – Short response (15%) ◦ assesses the application of a range of cognitions to several items, drawn from all Unit 3 topics. • Internal Assessment 3: Examination – Short response (15%) ◦ assesses the application of a range of cognitions to several items, drawn from Unit 4 Topics 1–5. • External Assessment: Examination – combination response (50%)	• Formulate ◦ statements of important assumptions and observations ◦ mathematical translation of important aspects of the task • Solve • use of mathematical knowledge for important aspects of the task • use of technology • a complete solution • Evaluate ◦ verified results ◦ statements about the reasonableness of the solution by considering the assumptions and observations ◦ statements of relevant strengths and limitations of the solution • Communicate ◦ use of appropriate mathematical language ◦ logical organisation of the response, which can be read independently of the task sheet ◦ justification of decisions using mathematical reasoning • Foundational knowledge and problem-solving ◦ consistently correct recall and use of mathematical knowledge ◦ authoritative and accurate communication of mathematical knowledge ◦ astute evaluation of the reasonableness of solutions ◦ use of mathematical reasoning to correctly justify procedures and decisions ◦ application of mathematical knowledge to solve problems in a comprehensive range of simple familiar, complex familiar and complex unfamiliar situations

Key Learning Areas (KLAs)	Sample subjects	Rationale	Instrument-specific marking guides (ISMGs)	Mark criterion and allocation

HASS (Humanities and Social Sciences)	Modern History
	
	Modern History 2025 v.1.2, (QCAA, 2024m)
		Through Modern History, students' curiosity and imagination are invigorated while their appreciation of civilisation is broadened and deepened. Students consider different perspectives and learn that **interpretations** and **explanations** of events and developments in the past are contestable and tentative.	• Internal Assessment 1 (Unit 1/3): Examination – Extended response (25%)
			  ◦ assesses the application of a range of cognitions to an unseen question.
			• Internal Assessment 2 (Unit 1/3): Investigation (25%)
			  ◦ an independent source investigation uses research and investigative practices to assess a range of cognitions in a particular context.
			• Internal Assessment 3 (Unit 2/4): Investigation (25%)
			  ◦ research a historical topic through the collection, analysis and synthesis of primary and secondary sources.
			  ◦ External Assessment: Examination – short response (25%)
				• Comprehending
				  ◦ thorough and mostly accurate use of terms in historical context
				  ◦ informed understanding of concepts related to the unseen question
				  ◦ detailed explanation of issues related to the unseen question
				• Devising and Conducting
				  ◦ development and application of a nuanced key inquiry question and relevant sub-questions
				  ◦ discerning selection of relevant evidence from primary and secondary historical sources
				  ◦ acknowledgment of different perspectives in the evidence from historical sources
				• Analysing
				  ° discerning selection and detailed examination of features of evidence from historical sources
				  ° Evaluating
				  ° discerning and well-reasoned judgements about the usefulness and/or reliability of evidence from historical sources in the stimulus
				• Synthesising
				  ° sophisticated historical argument that skilfully combines evidence from a range of historical sources to justify decisions
				• Communicating
				  ◦ conveys ideas related to the unseen question clearly and purposefully
				  ◦ all features of an essay in response to historical sources are consistently applied
				  ◦ minimal errors in spelling, grammar and punctuation

Key Learning Areas (KLAs)	Sample subjects	Rationale	Instrument-specific marking guides (ISMGs)	Mark criterion and allocation

HASS (Humanities and Social Sciences)	Business
	
	Business 2025 v.1.2, (QCAA, 2024e)
		Learning in Business integrates an inquiry approach with authentic case studies. Students become critical observers of business practices by applying an inquiry process in undertaking investigations of business situations. They use a variety of technological, communication and analytical tools to comprehend, analyse and interpret business data and information. Students evaluate strategies using business criteria that are flexible, adaptable and underpinned by communication, leadership, creativity and sophistication of thought.	• Internal Assessment 1: Examination – combination response (25%)
			  ◦ assesses the application of a range of cognitions to multiple provided items – questions, scenarios and problems.
			• Internal Assessment 2: Business report (25%)
			  ◦ research a specific problem through collection, analysis and synthesis of primary and secondary data.
			• Internal Assessment 3: Feasibility report (25%)
			  ◦ focuses on the analysis, interpretation and evaluation of ideas and information.
			• External Assessment: Examination – combination response (25%)
				• Describing
				  ◦ recognition of significant business facts and characteristics relating to strategic development
				  ◦ comprehensive description of a business situation relating to a business in the maturity stage of the business life cycle
				  ◦ comprehensive description of business environment/s that impact on marketing or operations
				• Explaining
				  ◦ identification of significant business concepts and strategies relating to marketing or operations for strategic development
				  ◦ comprehensive explanation of business concepts and strategies relating to marketing or operations for strategic development
				  ◦ purposeful use of business terminology and examples
				• Analysing
				  ◦ discerning selection of significant data and information relating to strategic development to analyse a business situation using the analytical tool
				  ◦ thorough interpretation of relationship/s and trend/s from the analysis
				  ◦ thorough implications for the business situation
				• Evaluating
				  ◦ proficient use of business criteria to make perceptive judgements for a marketing or operational strategy
				  ◦ thoroughly justified decision for the business situation
				  ◦ insightful recommendation/s for the business situation
				• Communicating
				  ◦ concise and logical sequencing and development of ideas to enhance meaning
				  ◦ features of the business report genre and use of referencing conventions are consistently demonstrated
				  ◦ minimal errors in spelling, grammar and punctuation

Key Learning Areas (KLAs)	Sample subjects	Rationale	Instrument-specific marking guides (ISMGs)	Mark criterion and allocation

Sciences	Biology
	
	Biology 2025 v.1.2, (QCAA, 2024d)
		Biology provides opportunities for students to engage with living systems. Students will learn valuable skills required for the scientific investigation of questions.
		Defining inquiry in science education
		As Abrams, Southerland, and Silva (2008, p. xv) stated in their book, Inquiry in the Classroom: Realities and opportunities: *Inquiry in the classroom can be conceived as a complex set of ideas, beliefs, skills and/or pedagogies. It is evident that attempting to select a singular definition of inquiry may be an insurmountable and fruitless task. Any single definition of inquiry in the classroom would necessarily reflect the thinking of a particular school of thought, at a particular moment in time, or a particular goal, and such a singular definition may serve to limit legitimate and necessary components of science learning.* ***However, operating without a firm understanding of the various forms of inquiry leaves science educators often 'talking past' one another, and often results in very muddled attempts in the classroom.***
			• Internal Assessment 1 (Unit 1/3): Data test (10%)
			  ◦ focuses on the application of a range of cognitions to multiple provided items.
			• Internal Assessment 2 (Unit 1/3): Student experiment (20%)
			  ◦ research a question or hypothesis through collection, analysis and synthesis of primary data.
			• Internal Assessment 3 (Unit 2/4): Research investigation (20%)
			  ◦ evaluate a claim by researching, analysing, and interpreting secondary evidence from scientific texts to form the basis for a justified conclusion about the claim.
			• External Assessment: Examination – combination response (50%)
				• Data test
				  ◦ consistent demonstration, across a range of scenarios, of
				  ◦ selection and correct application of scientific concepts, theories, models and systems to predict outcomes, behaviours and implications
				  ◦ correct calculation of quantities using algebraic, visual and graphical representations of scientific relationships and data
				  ◦ correct and appropriate use of analytical techniques to correctly identify trends, patterns, relationships, limitations and uncertainty
				  ◦ correct interpretation of evidence to draw valid conclusions
				• Forming
				  ◦ a considered rationale for the experiment
				  ◦ justified modifications to the methodology
				  ◦ a specific and relevant research question
				  ◦ a methodology that enables the collection of sufficient and relevant data
				  ◦ appropriate use of genre and referencing conventions
				• Finding
				  ◦ considered management of risks/ethical issues/environmental issues
				  ◦ collection of sufficient and relevant raw data
				  ◦ fluent and concise use of scientific language and representations
				• Analysing
				  ◦ correct and relevant processing of data
				  ◦ thorough identification of relevant trends/patterns/relationships
				  ◦ thorough and appropriate identification of the uncertainty and limitations of evidence
				• Interpreting and Evaluating
				  ◦ justified conclusion/s linked to the research question
				  ◦ justified discussion of the reliability and validity of the experimental process
				  ◦ suggested improvements and extensions to the experiment that are logically derived from the analysis of evidence

Key Learning Areas (KLAs)	Sample subjects	Rationale	Instrument-specific marking guides (ISMGs)	Mark criterion and allocation

Languages	Japanese
	
	Japanese 2025 v.1.2, (QCAA, 2024k)
		Additional language acquisition provides students with opportunities to reflect on their understanding of a language and the communities that use it, while also assisting in the effective negotiation of experiences and meaning across cultures and languages.	• Internal Assessment 1: Examination – short response (20%)
			• assesses subject matter from Unit 3 Topic 1: Roles and relationships.
			• Internal Assessment 2: Examination – extended response (25%)
			• assesses subject matter from both:
			  ◦ Unit 3 Topic 2: Socialising and connecting with my peers
			  ◦ Unit 3 Topic 3: Groups in society.
			• Internal Assessment 3: Multimodal presentation and interview (30%)
			• assesses subject matter selected from:
			  ◦ Unit 4 Topic 1: Finishing secondary school, plans and reflections
			  ◦ Unit 4 Topic 2: Responsibilities and moving on.
			• External Assessment: Examination – combination response (25%)
				• Analysing Japanese texts in English
				  ◦ draws a valid and well-substantiated conclusion about one of purpose/audience/context/tone from the stimulus
				  ◦ justifies with two pieces of supporting evidence from the stimulus
				• Analysing Japanese texts in Japanese
				  ◦ draws a valid and well-substantiated conclusion about one of purpose/audience/context/tone from the stimulus
				  ◦ justifies with two pieces of supporting evidence from the stimulus
				  ◦ conveys meaning in Japanese with minor errors
				• Exchanging information and ideas in Japanese
				  ◦ constructs meaning in response to stimulus by applying comprehensive knowledge of language elements in a conversation with minor errors in the use of
				  ◦ wide range of vocabulary
				  ◦ wide range of grammar
				  ◦ pronunciation and intonation
				  ◦ consistent register
				• Creating a multimodal presentation in Japanese
				• comprehension of ideas in three stimulus texts in Japanese related to the present and/or future choices

Key Learning Areas (KLAs)	Sample subjects	Rationale	Instrument-specific marking guides (ISMGs)	Mark criterion and allocation

Technologies	Design
	
	Design 2025 v.1.2, (QCAA, 2024f)
		The Design subject focuses on the application of design thinking to envisage creative products, services and environments. Designing is a complex and sophisticated form of problem-solving that uses divergent and convergent thinking approaches that can be practised and improved. Designers are separated from the constraints of production processes to allow them to appreciate and exploit innovative ideas.	• Internal Assessment 1: Design challenge (20%)
			  ◦ a supervised test that assesses the application of a range of cognitions to a provided design problem.
			• Internal Assessment 2: Project (30%)
			  ◦ focuses on a design process that requires the application of a range of cognitive, technical and creative skills and theoretical understandings.
			• Internal Assessment 3: Project (25%)
			  ◦ focuses on a design process that requires the application of a range of cognitive, technical and creative skills and theoretical understandings.
			• External Assessment: Examination – extended response (25%)
				• Exploring
				  ◦ insightful analysis of needs and wants
				  ◦ using relevant primary data about the stakeholder, secondary data about existing designs and designing with empathy knowledge
				  ◦ to identify the significant aesthetic, cultural, economic, social and technical features and constraints of design problems and the relationships between them
				  ◦ discerning description of the
				  ◦ features and constraints that define a HCD (human-centred design)
				  ◦ essential design criteria that integrate the stakeholder's specific requirements associated with the features and constraints of the problem and the principles of good design
				• Devising
				  ◦ range of credible ideas perceptively devised using designing with empathy in response to the HCD problem that demonstrate flexibility in ways of responding, originality, and detailed attributes
				• Evaluating and Proposing
				  ◦ critical evaluation of the strengths, limitations and implications of ideas against design criteria
				  ◦ discerning refinement of ideas based on the criteria and stakeholder information
				  ◦ coherent, logical and innovative HCD concept proposed, by combining attributes of multiple ideas and stakeholder information, in response to the design problem
				• Representing
				  ◦ sophisticated representation of ideas and a design concept using fluent sequences of ideation and/or schematic sketching to progress understanding in the develop phase

Key Learning Areas (KLAs)	Sample subjects	Rationale	Instrument-specific marking guides (ISMGs)	Mark criterion and allocation

HPE (Health and Physical Education)	Physical Education
	
	Physical Education 2025 v.1.2, (QCAA, 2024n)
		Physical Education fosters an appreciation of the values and knowledge within and across disciplines, and builds on students' capacities to be self-directed, work towards specific goals, develop positive behaviours and establish lifelong active engagement in a wide range of pathways beyond school. Students learn experientially through **three stages of an inquiry approach** to ascertain relationships between the scientific bases and the physical activity contexts.	• Internal Assessment 1: Project – folio (25%)
			  ◦ focuses on an inquiry process that requires the application of a range of cognitive and technical processes and skills, and theoretical understandings.
			• Internal Assessment 2: Investigation – report (25%)
			  ◦ research an ethical dilemma through collection, analysis and synthesis of primary data and secondary data.
			• Internal Assessment 3: Folio (25%)
			  ◦ focuses on an inquiry process that requires the application of a range of cognitive and technical processes and skills, and conceptual understandings.
			• External Assessment: Examination – combination response (25%)
				• Analysing
				  ◦ insightful analysis of relevant primary data and secondary data to ascertain the most significant relationships between the
				  ▫ demands of the specialised movement sequences and one movement strategy
				  ▫ relevant energy systems and fitness components
				  ▫ personal performance of the specialised movement sequences and one movement strategy
				  ▫ discerning synthesis of the specialised movement sequences, energy systems and fitness components and personal performance to devise a personal training strategy to optimise performance in the selected movement strategy
				• Evaluating
				  ◦ critical evaluation of the
				  ▫ effectiveness of the training strategy by appraising the outcomes and limitation
				  ▫ selected principles of training, training methods, energy systems and fitness components
				• Justifying
				  ◦ discerning justification, using primary and secondary data, of the
				  ▫ development of the training strategy to optimise performance
				  ▫ modification and maintenance of the training strategy to optimise performance
				• Communicating
				  ◦ appropriate decision-making about and accurate use of
				  ◦ at least two modes (visual, written, spoken) to achieve a particular purpose
				  ◦ language suitable for a technical audience
				  ◦ referencing and folio genre conventions
				• Demonstrating and Applying
				  ◦ effective demonstration of specialised movement sequences and two movement strategies (from two different principles of play for physical activities from the 'Invasion' or 'Net and court' categories) in authentic performance environments
				  ◦ effective application of quality of movement and one other body and movement concept to the performance of specialised movement sequences and two movement strategies (from two different principles of play for physical activities from the 'Invasion' or 'Net and court' categories) in authentic performance environments

Key Learning Areas (KLAs)	Sample subjects	Rationale	Instrument-specific marking guides (ISMGs)	Mark criterion and allocation

The Arts	Visual Art
	
	Visual Art 2025 v.1.2, (QCAA, 2024o)
		Students generate a focus for the body of work through **investigation of an individual inquiry** question in response to a specified stimulus.
		
		Inquiry learning
		Inquiry learning is fundamental to Visual Art. It requires students to solve problems through questions that have more than one possible resolution and emphasises the process of investigation when making and responding. Inquiry learning develops students' critical and creative thinking skills and contributes to their ability to participate in aesthetic processes. It encourages students to move beyond acquisition of facts to metacognition and developing understandings about concepts and focuses. It can also enhance personal and social skills by encouraging students to take responsibility for their own learning.
		
		The unit concepts and nominated contexts inform the teacher or student-directed focus and media as students make and respond, working as both artist and audience.
			• Internal Assessment 1 (Unit 1/3): Investigation – inquiry phase 1 (20%)
			  ◦ research a specific problem, question, issue, design challenge or hypothesis through collection, analysis, and synthesis of primary and/or secondary data.
			• Internal Assessment 2 (Unit 1/3): Project – inquiry phase 2 (25%):
			  ◦ focuses on a problem-solving process that requires the application of a range of cognitive, technical and creative skills and theoretical understandings.
			• Internal Assessment 3 (Unit 2/4): Project – inquiry phase 3 (30%):
			  ◦ implementation of ideas and representations informed by
			  ◦ constructed knowledge
			  ◦ an individualised inquiry question that directs targeted investigation
			• External Assessment: Examination – extended response (25%)
				• Developing – generating solutions
				  ◦ implementation of ideas and representations informed by
				  ▫ constructed knowledge
				  ▫ an individualised inquiry question that directs targeted investigation
				• Researching – reacting to stimulus
				  ◦ analysis and interpretation of the interrelationships between the individualised reaction to the stimulus and the artworks and/or practices of self and others
				  ◦ experimentation and exploitation of visual forms and media that demonstrate synthesis of new knowledge and understanding informed by the inquiry
				• Reflecting – considering ideas and information
				  ◦ evaluation of contemporary and/or historical influences and their impact on intellectual inquiry
				  ◦ justification of experimental approaches supported by critical understanding of the strengths, implications and limitations of the visual inquiry
				• Resolving – communicating as artist and audience
				  ◦ application of literacy skills through logical and connected ideas that unpack the inquiry question
				  ◦ realisation of responses that reach conclusions and express meaning about the interrelationships of knowledge and art practice to articulate a defined focus

In Confirmation, QCAA selects a sample of student responses from all endorsed Internal Assessments submitted for each general and general extension subject by every school. The sample is designed to represent a range of performance levels and is expected to have been moderated internally at the school prior to the submission of the provisional marks. For the selected sample, schools submit student responses and annotated ISMGs that indicate (a) the evidence identified, (b) the mark for each criterion on the ISMG, and (c) the overall mark assigned. Trained QCAA confirmers are allocated samples. They review the responses, provisional marks and evidence identified to evaluate the accuracy and consistency of the teacher judgements across the sample. If additional information is required before provisional marks can be confirmed, QCAA can request additional samples from the school. Once the Confirmation decision is reached, the result is communicated to the schools through the Confirmation app. Schools can request a review if a student's confirmed result differs from the provisional mark on one or more criteria and the school believes this is an anomaly or exception to the cohort pattern.

In 2024, approximately 165,000 samples across 46 general and general extension subjects were assessed by a casual workforce of more than 3,000 assessors. At the cohort level, more than 84% of Confirmation decisions aligned with schools' provisional marks in the last two cycles (QCAA, 2024c). Agreement rates vary by subject and technique; Figure 5.3 (QCAA, 2025c) presents mean agreement rates across all subjects and techniques. It should be noted that only two of the three internal summative assessments were administered in 2020 due to COVID-19 disruptions, and QCAA issued additional support materials at the time to help maintain consistency while students were learning from home. These figures may also differ from those in the QCAA Annual Report, as data collection and analysis may not be complete when the Annual Report is published.

Cohort agreement is lower than provisional agreement, as differences may affect only one criterion or performance level and thus not all students in a cohort. A cohort's provisional results can be confirmed at one level but not at another. Chapter 6 provides further information on the rate and direction of mark changes across subject areas, with full details in Appendix B.

Before 2023–24, the cohort Confirmation rate was steady at around 80%. The recent increase will be of interest to track in future cycles. In 2025, revised syllabuses and ISMGs will be implemented, which may influence consistency of judgements as teachers become familiar with the new marking guides.

Figure 5.3 Confirmation rates, 2020–2024

The Confirmation process is supported by a custom-designed software application with a clear workflow and internal messaging functions that enable communication between the confirmers, the Lead Confirmer for each team, and the Chief Confirmer. The app allows direct access to the Lead Confirmer, who can respond to requests for support and undertake calibration as needed. This helps ensure consistent understanding of performance-level requirements and accuracy in review decisions. When a school's provisional mark for a criterion is not supported, the Confirmer identifies the descriptors on the ISMG that better align with the student work, records the alternative mark and cites the key evidence supporting the decision in the app. The system also allows confirmers to flag instruments with unresolved issues for review by the Lead or Chief Confirmer.

Qualitative evidence from confirmers was gathered through surveys (2,064 surveys distributed, 610 responses) and free-response comments (284). Overall, they expressed positive views across a range of topics. However, some concerns were noted, particularly regarding the availability of sufficient confirmers to sustain the system, workload and a perceived lack of feedback on their work. Time pressures were a recurring theme, particularly among Lead Confirmers. Further detail on survey responses is provided in Appendix E.

Findings

- Confirmation assures the consistent use of the ISMGs across endorsed assessments to assign marks to student responses
- Cohort Confirmation rates have been slowly increasing over time
- Provisional mark agreement rates are very high, above 90%, and have been increasing over time

Recommendations

- QCAA to analyse Confirmation rates within and across subjects for each validity and accessibility priority and examine the results for informative patterns

Suggested Operational Adjustments

- QCAA to publish and explain rates for each cycle of cohort and provisional mark Confirmation
- QCAA to review the Confirmation schedule and evaluate the feasibility, schedule impact, and financial cost of a longer review period as well as workload of Lead and Chief Confirmers to ensure that sufficient time is allowed for planned training, operational activity and quality assurance tasks
- QCAA to design and implement a system for providing quality assurance data and feedback to all confirmers, to collect additional appropriate data to support such feedback , and assure that Lead and Chief Assessors meet established minimum feedback expectations

5.4 External Assessment

External Assessment (EA) is the final assessment in general and general extension subjects and is the only assessment in general (senior external examination, or SEE) subjects. The QCAA website provides clear information on EA processes, content and development but does not explicitly state the purpose or the assessment attributes EA is designed to address. From the Standards (AERA et al., 2014) it can be inferred that EA supports three foundational areas: validity, reliability/precision of measurement and fairness in testing.

EA plays a central role in the validity of the QCE by ensuring accurate construct representation and comparability of results for students within a course. Because the EA is administered at the same time and under standardised conditions for all students in a subject cohort, QCAA has maximum control over these areas compared with more limited control in the IAs. The difference in control is supported by evidence from correlations between student ability estimates from the IAs and EA. These correlations are moderate to high, indicating that the instruments measure the same construct overall, but with some variability in outcomes. Through a carefully controlled development and marking process, the EAs demonstrate high reliability over time with low measurement error. Further analysis is available in Appendix B. Two inferences might be supported by evidence from the EAs:

- Concurrent evidence – EA results contribute to the overall subject result and credit. They support the inference that the student has met or exceeded the standard required to earn a QCE qualification.
- Predictive evidence – This refers to whether assessment outcomes relate to or predict later performance on similar criteria. QCAA does not currently collect or analyse longitudinal predictive data of this type.

According to QCAA (undated–b), EAs have high accessibility. Teachers and students are provided with common syllabi that describe the specifications, conditions and assessment objectives of the summative EA. This ensures students have a clear understanding of expectations for demonstrating their learning. From the perspective of the *Standards* (AERA et al., 2014) and the broader conception of fairness, the procedures appear fair: all students undertake the assessment under standardised conditions set by QCAA. Psychometric analysis also shows minimal differential item function (DIF) across sampled subjects. DIF is a statistical procedure used to identify test items that function differently for different groups of test-takers after the groups are matched on overall ability. In simple terms, DIF flags items that may unfairly favour one group over another. The DIF results indicate little or no measurement bias in the EA results.

External Assessments are conducted in all schools with students enrolled in Units 3 and 4 in the relevant general subjects. Content is determined by QCAA through the subject syllabus. EAs are developed and marked by QCAA but administered by schools under standardised conditions on the same day and time across the state. Students' achievement in the EA, together with results from the IAs, contribute to the final subject result. EAs are not used for Essential English and Essential Mathematics; instead, these subjects use common Internal Assessments (CIAs) developed by QCAA and marked by schools. Applied subjects do not have externally developed exams.

The first operational administration of EAs in the new QCE, in 2020, covered 46 general and general (extension) subjects and 22 SEEs and was undertaken by around 37,300 students across Queensland. Approximately 222,000 response papers were marked by about 3,000 assessors. Since then student numbers have remained steady at about 38,000 per year, with 230,000 to 240,000 papers, requiring marking by around 3,000 markers annually.

Subject reports are published early in the year following the operational delivery of the assessment, providing analysis of the IA and EA outcomes, annotated samples of student work against the External Assessment Marking Guide (EAMG), reflections on strengths in IA submissions and identification of areas for improvement.

The number of SEE subjects has decreased, from 22 in 2020 to 13 in 2023. Language SEEs were always intended to continue, as they typically attract small candidate numbers and resource demands can be shared with another jurisdiction. Only two non-language SEEs remain: English and General Mathematics. These subjects are designed for learners reengaging with formal education to demonstrate the literacy and numeracy levels required for a QCE without school enrolment.

External Assessment development

Development of high-stakes external examinations is a complex process in any educational system. It generally follows a standard set of stages (Lane et al., 2015) and requires strict adherence to a complex schedule. This involves collaboration with external content experts and agency staff, while maintaining high levels of security to prevent unintentional exposure of secure test materials. Such processes can falter for many reasons (Constantinou, 2024), and any errors in high-profile assessment programs attract intense public scrutiny.

A recent Australian example is the series of errors in Victoria's VCE exams over three years from 2022 to 2024. The 2024 errors were particularly serious, with secure test materials exposed publicly. Reports suggest this occurred due to rushed implementation of a new and insufficiently verified automation process. In response, the Victorian Government ordered a 'root-and-branch' review (Blacher, 2025) of the Victorian Curriculum and Assessment Authority (VCAA). The review examined delivery procedures and recommended improvements in governance, staffing, resourcing and operational culture within the VCAA to restore public confidence in the examination process and validity of inferences based on the results.

Developing Queensland's EAs is a complex process with a series of built-in quality assurance steps. There are six learning areas (sciences, mathematics, HPE/technology, humanities, and social sciences, English and languages, and the arts) each with a Learning Area Manager (LAM) and Principal Education Officers (PEOs) for every subject. QCAA engages project officers, academics and subject-matter experts in this process and defines clear structures for collaboration, illustrating a broad picture of effectiveness. The process is depicted in Figure 5.4 overleaf.

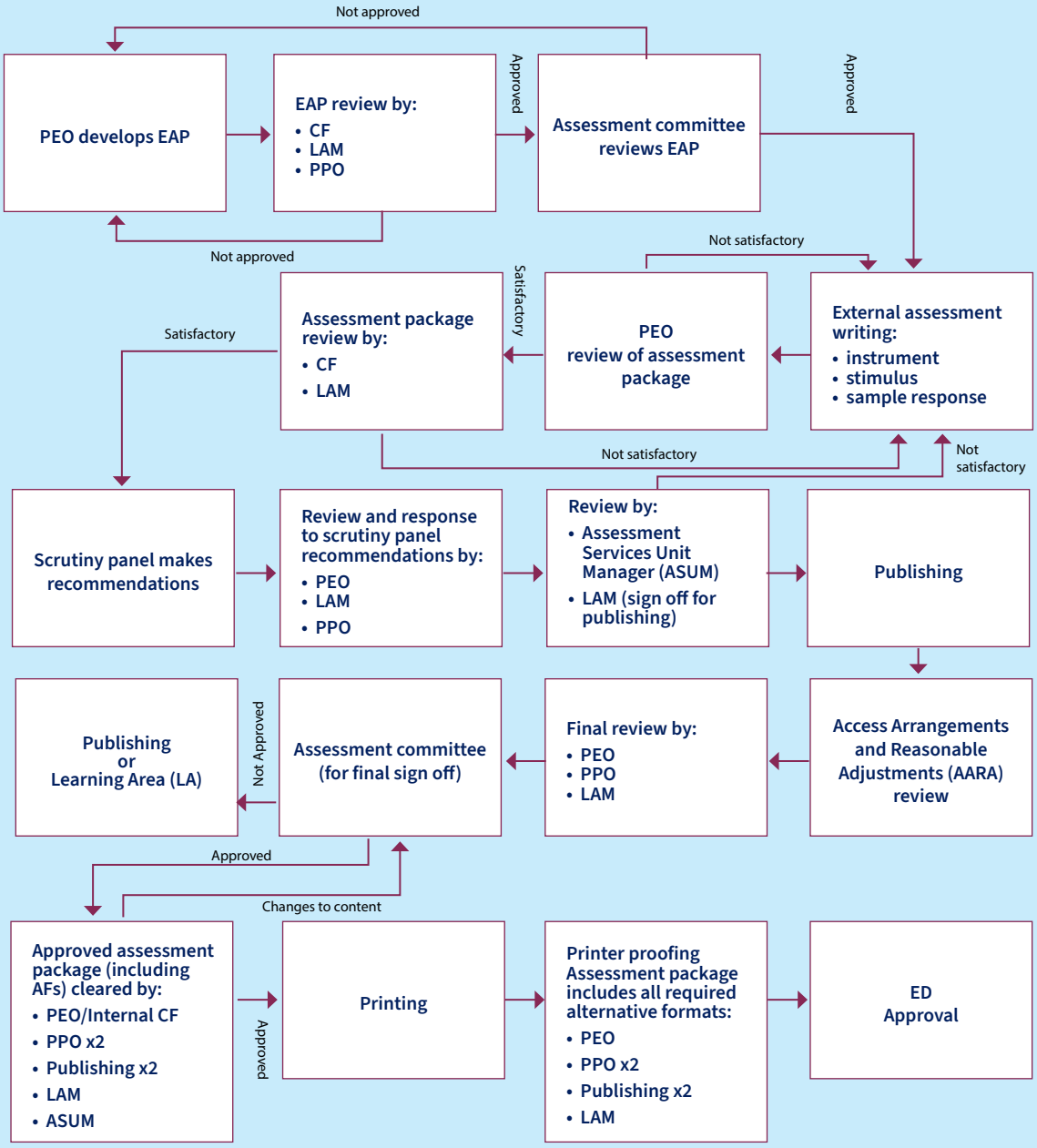


Figure 5.4 Flow chart of EA development

The broad outlines of the QCE EA and EAMG development process parallel those of some of the most highly regarded assessment programs in the world, such as PISA (OECD, 2024), NAEP (National Center for Education Statistics, undated), and NAPLAN (ACARA, undated-b). Differences arise in administration mode and translation requirements, which involve different procedural and quality control steps. Each year, EA development and marking are informed by a subject-specific technical report. These reports provide feedback to the assessment development team, supporting integration of technical findings into ongoing quality improvements. This feedback loop is a key feature of high-quality assessment design and development.

Qualitative evidence from the markers was gathered through surveys (3,121 surveys sent, 1,028 responses) and free-response comments (498 comments). As with Endorsement and Confirmation, survey responses were generally positive about the process, timing and security of marking. Markers typically reported that the expected pace and duration of their work was achievable and expressed confidence in the accuracy of their marking. In the free-response comments, the most common themes were high workload and low pay rates, along with requests for more timely and concrete feedback on their work. More detail on the survey responses is available in Appendix E.

Findings

- The QCE EA and EAMG development process are high quality, well controlled and parallel some of the most highly regarded assessment programs in the world
- There are several well-defined roles in the development process with clear and consistent documentation of the responsibilities of each role
- The EAs generally have high reliability and limited error of measurement
- The correlation between the IA and EA results are moderate to high across subjects, indicating consistency in the underlying construct measured

Recommendations

- QCAA to clearly define the EA's intended purpose and the attributes of quality assessment that EA will address on the publicly available QCAA website
- QCAA to incorporate checks to assure adherence to EA development process as designed and monitor for schedule slippage
- QCAA to review the marking schedule and evaluate the feasibility and schedule impact of a longer marking period to ensure that sufficient time is allowed for planned training, operational activity and quality assurance tasks
- QCAA to design and implement a system for providing quality assurance data and feedback to all markers, and ensure that Lead and Chief Assessors meet the pre-determined minimum feedback expectations

5.5 Applied subjects quality assurance

For Units 1–2 of applied, applied (essential), short courses and general subjects, assessment is not submitted for Endorsement or Confirmation. Units 3–4 are endorsed annually in Essential English and Essential Mathematics only.

For all applied subjects, schools must maintain a quality management system and complete annual reviews of assessment and judgements. Beginning in 2025, applied subjects other than applied (essential) subjects are scheduled for quality assurance on a rotating basis. If a school has students enrolled in Units 3–4 for a scheduled subject, it must participate in an applied subject quality assurance meeting.

For applied subjects, schools are required to submit two summative IA instruments for one selected topic in Unit 3, along with two marked samples of student responses for each instrument. Samples should represent high- and mid-range performance. Each marked sample must be accompanied by an annotated instrument-specific standard showing how the judgement was made.

For applied (essential) subjects, schools are required to submit the endorsed IA1 and three marked samples of student responses representing high, mid-range and low performance. They must also submit the CIA for Unit 3 and three marked samples of student responses, with annotated marking guides indicating the evidence for each judgement.

At the applied quality assurance meeting, the assessments and student responses are reviewed. This process ensures the quality, comparability and reliability of schools' implementation of these IAs. The process is supported by a custom Applied QA app, used to review submissions during meetings and provide feedback on assessment design, accuracy of the teacher judgements and application of instrument-specific standards and marking guides. Evidence includes matching student responses to relevant criteria as well as the overall grade assigned.

If the reviews find that policies and procedures have not been applied as required, QCAA provides advice and intervention support. Schools are expected to apply the advice to remaining assessments for the current cohort and to strengthen future practice, although there is no mandated verification of incorporation of feedback as there is in Endorsement. Feedback on the accuracy and reliability of marks for the sampled students may also apply to the broader cohort and is expected to be addressed before results are entered into the Student Management app.

Applied quality assurance procedures are a lighter version of those for general subjects. Only a subset of applied subjects is reviewed annually, while all general subjects undergo annual quality assurance. QCAA provides advice on applied subject submissions and identifies schools that may need further support. For general subjects, which follow the Endorsement-

Confirmation-External Assessment cycle, schools cannot proceed if submissions are unsuccessful at any stage. Given the large number of students enrolled in applied subjects, closer monitoring may be warranted.

Applied subjects in Queensland are designed for students seeking pathways into vocational education, training or employment. All applied subjects include development of literacy, numeracy and 21st-century skills. Four unit applied subjects may contribute four credits towards a QCE, but only one can contribute to an ATAR, and it must be combined with four general subjects. The above quality assurance approach and the restrictions on ATAR contribution create a perception that applied subjects are less valued than the more 'academic' general subjects, which are oriented towards university pathways. This reflects a wider divide in public perception of applied/vocational and academic education, a pattern seen internationally (Gore et al., 2017; Lee et al., 2022; Reynolds, 2004).

Findings

- QCAA has implemented annual quality assurance procedures for a sample of Applied and Applied (Essential) subjects
- QCAA provides advice support and intervention to schools whose assessments are not fully aligned with the specified policies and procedures
- QCAA does not provide direct follow-up on advice given as part of applied quality assurance, and schools are not required to provide evidence of assessment modification based on such advice

Recommendations

- QCAA to consider ways to strengthen quality assurance procedures for Applied subjects

5.6 Enhancements: Assessment literacy modules and educator training

The evaluation team was given access to the QCAA portal, including the assessment literacy training modules and the Learning Hub. In addition, the team attended an online training session for new confirmers.

Assessment literacy modules

QCAA invests in developing and supporting teacher assessment literacy as part of its professional learning and development systems. There are seven modules in assessment literacy. The first three cover:

1. Attributes of quality assessment.
2. Developing valid and accessible assessment.
3. Making reliable judgements

There are four additional modules specialised for educators who wish to further develop their knowledge and/or apply for assessor roles within the QCAA system. These cover:

1. Endorsement.
2. Confirmation.
3. External Assessment writing.
4. External Assessment marking

All modules follow a similar pattern, with sections of content and explanation supported by frequent examples. These are followed by questions to assess knowledge, skills and understanding. The academic content area addressed in each module items varies, but knowledge of the specific content area is not necessary to respond. The questions include a variety of formats which can be automatically scored, including multiple choice, card-sort items, matching and simple short-response items. Generally the quiz can be repeated if the required accuracy is not achieved, and 100% accuracy may be needed to progress.

The modules are generally well constructed, and the user interface is intuitive. However, some definitions of key concepts are not well aligned with current thinking in the field of assessment or standard practices. For example:

- validity framed as a property of an assessment is no longer broadly accepted in the field
- accessibility is typically included as part of fairness in professional assessment frameworks
- the module's definition of accessibility is not aligned with that used in the field of assessment
- reliability is more appropriately considered as a component of validity and not a separate attribute
- discussion of reliability should include concepts related to errors/precision of measurement and the impact on drawing valid inferences
- intra- and inter-marker accuracy are measured by agreement statistics, not reliability statistics
- inter-marker agreement is not a measure of accuracy
- attributes of items – such as defining selected-response items as only multiple choice or classifying short-response items into categories – are unnecessary and unduly restrictive in terms of the characteristics of quality assessment items

Some content included in the modules, such as the validity and reliability priorities and their explanations, are not available on the public website.

Attributes of quality assessment

The QCAA website (2021b) lists the attributes of quality assessment (AQA) as validity, accessibility and reliability. The Standards (AERA et al., 2014) serve as the international touchstone for assessment and testing in education and psychology. The Standards list the foundational properties of high-quality assessments as validity, reliability/precision and errors of measurement, and fairness in testing. The Standards group access, accuracy and appropriate score use under the broader concept of fairness (AERA et al., 2014). QCAA's characteristics are mostly well aligned to those in the Standards. Exceptions include the narrower focus on accessibility in contrast to fairness and the lack of training content on precision and errors of measurement.

Darling-Hammond et al. (2013) list five major elements that describe assessments capable of evaluating deeper learning:

1. Assessment of higher-order cognitive skills.
2. High-fidelity assessment of critical abilities.
3. Standards that are internationally benchmarked.
4. Use of items that are instructionally sensitive and educationally valuable.
5. Assessments that are valid, reliable and fair

The QCAA attributes of quality assessment are aligned to element 5, and other sections of this evaluation address additional elements from the list.

Koretz (2008) provides additional information regarding the use of the terms 'valid' and 'reliable'. Validity is a frequently misused term; it is not a property of an assessment or a test score, although it is often applied that way. Validity is a property of the desired inference based on the measurement result. Note that this is at variance with the QCAA definition, in which validity is a property attributed to the assessment, not the inference drawn from the results. Validity is best conceived as a unitary concept. It should be supported by a range of evidence sources with the goal that the set of collected evidence supports the intended interpretation or use of the assessment result. The definition on the QCAA website (2021b) names two aspects of validity (content and construct), as well as authenticity of tasks, realism of context, alignment with curriculum and demonstration of performance levels. These are all important aspects of validity.

Validity can be compromised in three important ways (Koretz, 2008).

1. Construct under-representation – For an inference about a student's ability in a content domain, the domain must be adequately sampled by the assessment
2. Construct-irrelevant variance – To isolate the skill about which an inference is desired, the content of the measurement must be appropriately focused and not overly broad. Factors that are extraneous to the target skills should be minimised or eliminated
3. Malpractice – There are a range of behaviours that can impact validity of inferences, including cheating, gaming and narrowing the curriculum to the point that teaching targets only the domain sample that will be tested

Reliability is a measure of consistency in assessment. Reliable scores provide the same (within measurement error) score for a candidate over repeated cycles of measurement, assuming no learning occurs from the measurement itself. For assessments requiring constructed responses, reliability also means that a response should receive the same mark regardless of the individual marker assigned (McCaffrey et al., 2022). The reliability of the assessment has a direct impact on the strength of evidence that it can offer to support valid inferences. However, reliable assessments may still be overly broad or under-represent a construct, and markers can agree on a result (i.e., be consistent) but still be incorrect or inaccurate. QCAA assures reliability through consistent curricula and syllabi, common marking guides and moderation processes. Together, these provide a type of alternate form replication. In this process, sets of assessment items are developed to be similar enough to serve as different versions of the same test. Alternative forms can then be used as a measure of reliability.

The QCE system includes controls to address these factors. Construct representation is bounded by the common curriculum, syllabus and marking guides used in each course. Malpractice is addressed by the use of an externally set and marked assessment, as well as school-based internal moderation. Ratification procedures are in place to identify anomalous outcomes and work with schools to address them.

Accessibility is the third characteristic of quality assessment mentioned by QCAA (2021b). On its website, QCAA defines accessibility as "the extent to which the assessment provides all students with a clear understanding of how to demonstrate their learning". More detailed information is included in dot points, focused mainly on content exposure, information about assessment structure, student feedback and clear criteria. The explanatory text also refers to inclusive strategies, access and adjustments.

To improve accessibility and prepare students for the new EAs, QCAA developed sample papers in more than 50 subjects. These papers allowed schools to prepare students for EAs. Seeing models of future assessment is important for students learning and helps reduce unnecessary stress and anxiety (AERA et al., 2014). Other countries provide such resources to ensure transparency and support student understanding of the

assessment system (e.g., the Irish Leaving Certificate – Baird et at., 2014). Transparency and sample papers are also important for building public trust. The QCAA and QCIA policy and procedures handbook (2025) outlines and emphasises practices which are transparent for students, parents, guardians and the public.

QCAA's definition of accessibility varies somewhat from the definition commonly used in the field of assessment. For example, the Standards (AERA et al., 2014) define accessibility as

> the notion that all test takers should have an unobstructed opportunity to demonstrate their standing on the construct(s) being measured … access to the construct the test is measuring can be impeded by characteristics and/or skills that are unrelated to the intended construct and thereby can limit the validity of score interpretations for intended uses for certain individuals and/or subgroups in the intended test-taking population. (AERA et al., 2014, pp. 49–50)

Accessibility is presented in the Standards for Educational and Psychological Testing as one component of the broader construct of fairness:

> [F]airness to all individuals in the intended population of test-takers is an overriding, foundational concern, and that common principles apply in responding to test-taker characteristics that could interfere with the validity of test score interpretation … the equivalence of the construct being assessed is a central issue in fairness. (AERA et al., 2014, p. 50)

There is a qualitative difference between a student being given a clear understanding of how to demonstrate their learning and a student being given an unobstructed opportunity to do so.

Fairness is broadly framed as "responsiveness to individual characteristics and testing contexts so that test scores will yield valid interpretations for intended uses … a fair test does not advantage or disadvantage some individuals because of characteristics irrelevant to the intended construct" (AERA et al., 2014, p. 50). QCAA's definition of accessibility is narrower and does not explicitly reference concepts from universal design.

Universal design

Fairness in the sense of maximising access to assessment for all students, can be enhanced through the application of universal design of assessment (UDA) systems in the EA development process. Within the QCAA portal, a non-mandatory Learning Hub module titled "Providing Accessible Assessment Tasks" introduces teachers to principles from universal design for learning (UDL).

UDL (CAST, 2024) is a curriculum design framework aimed at ensuring all learners have equal opportunities to succeed and demonstrate capability. UDL is not deficit-based; instead, it reflects a view that barriers to learning lie in the curriculum rather than in the learner. Its principles guide the creation of flexible learning experiences that address diverse learning needs and preferences. In the QCAA Learning Hub, there are P–6 and Years 7–12 courses on providing accessible assessment tasks. These are available to educators with a QCAA portal login, but they are not part of the assessment literacy modules. Despite their focus on assessment tasks, the content of these courses is explicitly about universal design for *learning*. While UDL is a key part of equitable education, the CAST framework is oriented primarily towards classroom instruction, with a smaller focus on assessment. It is also aimed at the needs of the learner, without considering other actors and elements in the education system.

UDA (Thompson et al., 2002) extends UDL principles to assessments, particularly large-scale ones, and tends to encourage a checklist approach at the item level for compliance. While valuable, UDA is too narrow to encompass the full range of assessment types and approaches used in education. UDA systems (UDAS) – (Laitusis & Karvonen, 2025) addresses these limitations, emphasising test development that maximises usability for all examinees in the target population, regardless of their specific profile. The UDAS framework is shown in Figure 5.5 (adapted from Laitusis and Karvonen (2025).

Figure 5.5 UDAS principles and guidelines

Findings

- QCAA provides extensive assessment literacy training through 3 core and 4 specialised modules, plus additional professional development materials in the Learning Hub. These modules have a range of strengths:
  - Assessment literacy training is easily accessible to all Queensland educators through the QCAA portal
  - The training provided for specialised assessors is engaging, practical, valuable and well aligned to the operational requirements of assessor roles

- There are opportunities to improve the assessment literacy training:
  - Assessment literacy training is not mandatory for Queensland educators although the core modules are recommended for all educators
  - The content of the assessment literacy training modules is not well aligned with current thinking in the field of assessment
  - The limited increase in endorsement and confirmation rates over time does not provide convincing evidence to support inferences about the impact of the current assessment literacy training for educators on their enacted assessment development practices
  - The content of the assessment literacy training modules is neither adequately aligned to educators' individual assessment practices nor sufficiently practical to have a meaningful impact on educator's assessment practices

- Incorporation of a module on universal design of assessment systems into the core assessment literacy modules would enhance their value:
  - The course including content on universal design for learning (UDL) in the QCAA Learning Hub is not part of the core assessment literacy training
  - The UDL course content is focused on universal design for learning but is titled "Providing accessible assessment tasks"

Recommendations

- QCAA to update and revise the assessment literacy training modules to align with contemporary research
- QCAA to collaborate with school authorities to encourage all Queensland senior secondary teachers to successfully complete the core assessment literacy training modules within 2 years of employment

Suggested operational adjustments

- QCAA revise or redevelop the assessment literacy courses as follows:
  - Align the 'attributes of quality assessment' module with the most recent edition of the AERA/APA/NCME Standards, so that attributes of quality assessment are redefined as validity, reliability/precision and errors of measurement, and fairness in testing
  - Align the universal design for learning course with the universal design for assessment systems (UDAS) framework and include it in the core assessment literacy training modules
  - Develop more practical exercises and individualised content relevant to the educator sitting them, extending beyond the current low-level assessment of knowledge and skills
  - Include immediate automated feedback on incorrect responses and provide rationales for all correct and incorrect responses to quiz items in the modules
- QCAA promote the changes and review their impact:
  - Conduct a public awareness campaign to highlight changes in the assessment literacy modules
  - If the recommended change affects operational procedures, include a description of the impact on the publicly accessible website and in the public awareness campaign
  - Make materials defining validity and reliability priorities and explanations of them available on the public QCAA website

5.7 Inter-relationship between assessment processes and overall structure

The QCE system delivers a unique approach to quality assurance of secondary education. Like any complex system there is room for improvement, but the system is functional, sound and delivers quality results.

The three primary processes QCAA has put into place to monitor and assure assessment quality are Endorsement, Confirmation and External Assessment. These stages are designed to address the three key attributes of quality assessment in the QCE system: validity, accessibility and reliability. In this section, we use these attributes as described by the QCAA rather than the broader foundational properties as defined in the Standards (AERA et al., 2014). Those distinctions have been discussed previously in the sections on the QCE QA processes and will not be repeated here.

Table 5.3 below is based on the professional judgment of the evaluators in assessing the extent to which each major process of the QCE assessment structure contributes evidence to the attributes of quality assessment as defined by QCAA. The reasoning for each conclusion is briefly described in the column below the indicator. This table is not intended to be definitive about the quality or quantity of evidence in each case, as professional opinions may differ and data collection may evolve over time. Rather, the goal is to illustrate the overall QCE evidence profile based on the evidence collected in this project.

As noted previously, it is not expected that a single assessment provides high levels of evidence for all attributes of quality assessment. Instead, the goal of the system is to ensure that there are high levels of evidence provided *across* the set of components. The pattern in the table shows the inter-relationships between the processes in terms of the evidence each provides and how that fits together into a comprehensive evidential validity argument. Overall, the profile indicates that the QCE assessment structure does provide high levels of evidence to support valid inferences about student achievement and proficiency.

Table 5.3 Cross-tabulation of QCE QA processes and AQA evidence

Process / Attribute of quality assessment	Validity evidence	Accessibility evidence	Reliability evidence
Endorsement	**Construct: Moderate** **Predictive: Limited** **Contextual: High** Endorsement provides a moderate level of validity evidence for content and construct. The IAs are endorsed if they meet the criteria for validity in the quality assurance tool for that subject. Since the instruments have natural variability in their contextualisation and construction, there is an associated limit on the contribution to the evidence for construct validity. There is some evidence of predictive power of the confirmed responses to the IAs for the EA in the moderate correlation between the estimates of student ability calculated in the psychometric analyses. The IAs are developed by educators in the local context and culture and thus provide solid evidence for validity of use in the target population of students.	**High** Endorsement assures that IAs are aligned to the syllabus and cover the content of the course. This central review of the IAs assures that teachers have actioned the accessibility requirements for assessments as set out by the QCAA, in addition to verifying the content evidence. Endorsement cannot verify the instructional practices that are described as part of accessibility. Those aspects of accessibility should be monitored and evaluated by the local school authorities, and the results should be evidenced in the assessment submissions to Endorsement.	**Moderate** IAs are developed in alignment with the prescribed curriculum, syllabus and ISMGs. Each of these control mechanisms increases the reliability of the assessments submitted to Endorsement by reducing construct-irrelevant variance and ensuring appropriate construct coverage. Endorsement assures that the state-wide set of Internal Assessments deliver a version of alternative-form replication. Each endorsed IA is an alternative form that assesses the construct defined in the syllabus and each IA response is evaluated against the same ISMG. This consistency will increase the overall reliability of the IA results.
Confirmation	**Construct: Moderate** **Predictive: None** **Contextual: Moderate** Confirmation provides a moderate level of validity evidence for content and construct, like Endorsement. As the ISMGs are specified for use across the endorsed assessments, they impose a level of control over the content that support valid inferences that the student responses are aligned with the syllabus and course expectations. As with Endorsement, Confirmation of accurate marking student responses to assessment items against the ISMG to items that were developed to be contextually and culturally appropriate. The contextualisation is bounded by the use of a common ISMG.	**Limited** Confirmation serves to assure that marking of student responses is consistent and is aligned with the common ISMG. The accessibility of the IA has largely been established through classroom instructional practices prior to assessment administration and through IA design and development. Confirmation can contribute some evidence about the success of accessibility in the prior stages through detection of issues such as student misunderstandings of items based on their responses, speediness of IAs based on student response completion, and the use and success of reasonable adjustments.	**High** The concept of reliability is focused on whether the endorsed IAs will produce a consistent, comparable result. Confirmation is targeted towards providing evidence on that point. The process of determining if the ISMG has been used consistently to mark student responses across the alternate forms of the IA gives direct, clear evidence of reliability. In addition to formal QCAA Confirmation, educators engage in internal moderation conversations with colleagues to assure marking consistency, providing additional evidence to support inferences around IA reliability.

Process / Attribute of quality assessment	Validity evidence	Accessibility evidence	Reliability evidence
External Assessment	**Construct: High** **Predictive: None** **Contextual: Limited** Because EAs are tightly controlled and developed to be aligned the content and curriculum of the subject, they provide robust evidence for the content and construct aspects of validity. This is also supported by the psychometric analyses in which only small amounts of differential item functioning were found, and that model fit was good. As noted previously, EA contributes concurrent evidence to support inferences about performance in the subject and the award of a QCE. The central development and administration controls limit the extent to which the EA can be adjusted for local culture or context.	**Limited** The structure and administrative conditions of the EA place boundaries on how much accessibility can be provided by teachers to students and still maintain acceptable levels of standardisation. There is appropriately limited information on the content of the assessment, although exemplar examination content is available to give guidance on administration conditions, timing, response format and item types. QCAA provides a range of accommodations through the formal Access Arrangements and Reasonable Adjustments (AARA) procedures or through approval by a school administrator. Reasonable adjustments can be provided through a formal application process to either school administrators or the QCAA.	**High** EA contributes strongly to the reliability of the QCE system results. EAs are developed centrally by QCAA, with standardised administration conditions. EA markers are employed, trained and monitored by QCAA as well. EA has the highest level of control of the three processes, and as a result, is the most reliable and has the least measurement error and variability.

As the matrix above illustrates, no single process in the QCE assessment system provide high levels of evidence for all three attributes of quality assessment. Taken together, however, the processes provide high levels of evidence across the attributes. The complete set of evidence offers strong support inferences about the quality and utility of a QCE awarded based in part on the evidence of the new assessment system.

Students benefit from this inference, as the value in the tertiary education, training and employment market of the QCE is enhanced by the quality of the supporting evidence. Students also benefit from the quality assurance controls in the system. Endorsement is intended to reduce or eliminate bias and inconsistency in the evaluation of student learning and ability, ensuring that students have the opportunity to learn and to demonstrate their learning equally regardless of their location, demographics or other extraneous factors. Endorsement allows room for adjustment for local culture and context while maintaining standards for quality. Confirmation also reduces potential bias, including in how student responses are evaluated. Individual teacher preferences are controlled through assurance that assessment responses are marked fairly and consistently against the ISMG and across the IAs. External Assessment, the most tightly controlled element of the processes, increases public and policymaker confidence in the comparability and consistency of the QCE outcomes through tightly controlled processes for content, development procedures and administration conditions.

The QCE has a weighted balance of Internal and External Assessments. The Matters and Masters review (2014) recommended that all subjects have a final result weighted 50% to the IAs and 50% to the EAs. The Queensland Government partially adopted this recommendation, choosing to apply the 50/50 weighting in mathematics and science but using 75% IAs and 25% EAs for other subject areas. The adoption of a weighting scheme to combine assessment results is primarily a subjective choice, driven by the desired inferences to be made using the evidence from the results. Responses to surveys as part of this evaluation indicate that the current weighting scheme is not without controversy.

There is extensive technical literature on the weighting of items, sub-scores and test scores going back more than 50 years (see, for example, (Ganzfried & Yusuf, 2018; Sinharay, 2019; Stanley & Wang, 1970; Wang, 1998)). Approaches range from simple procedures to complex statistical techniques. Schools are required to implement authentication processes to assure the source of student work and use design principles around authentic assessment to create classroom assessments and their IAs. This inevitably produces some level of variability in the IA results. From a purely psychometric perspective, weighting would be chosen to optimise the statistical reliability of the resulting score. This would suggest that the EA, with its tightly controlled processes for development, administration and marking, should carry more weight than the more variable IAs. The EA is also likely to be best insulated from the inappropriate use of generative Artificial Intelligence (GenAI) tools and other violations of academic integrity. However, the strictly technical viewpoint is just that: one option among many.

The optimal weighting of the QCE component assessments would best be determined through a facilitated discussion to agree the intended use(s) of the combined scores. Once this has been agreed, technically qualified personnel can calculate the impact of a range of approaches for consideration and a decision as to which one best supports the intended use(s).

Findings

- The QCE system delivers a unique approach to quality assurance of secondary education with a functional and sound system that delivers quality results
- Across all processes there are high levels of evidence of each of the attributes of quality assessment provided from one or more of the procedures, supporting its fitness for the current purpose
- The weighting of IA and EA differs across content areas in line with Queensland policy

Recommendations

- QCAA to clarify rationale and purpose of differential weighting of Internal and External Assessments and select principled and consistent approach across subjects

Suggested Operational Adjustments

- QCAA to consider commisioning studies providing increased predictive validity evidence for attainment of the QCE in its most common uses, such as university admission, vocational training or employment

5.8 Enhancements: Technology and GenAI operational tools

QCAA has built a unique assessment system to support the award of the QCE. The cycle of Endorsement, Confirmation and External Assessment provides a strong evidence base for inferences about student achievement and capability. However, the system depends substantially on casual staff to complete these processes. During the evaluation, stakeholders (including QCAA staff) highlighted potential challenges with the long-term sustainability of this approach, particularly considering teacher availability and the limited number of instructors in some subject areas in Queensland.

New and emerging technology tools could help reduce the system's reliance on casual assessors. Some potential options are outlined here within the processes of Endorsement, Confirmation and External Assessment. These represent only one possible direction for technology integration.

Endorsement

Endorsement is the process through which educators submit their summative Internal Assessments for Units 3–4 in general and general extension courses. These assessments must align closely with the syllabus documents and be marked against a common instrument-specific marking guide (ISMG). Submitted assessments are reviewed by QCAA for alignment with these documents and for adherence to the attributes of quality assessment.

IAs are tightly controlled: they must use prescribed cognitive verbs, mark allocations and task types. This strict guidance is both an asset and a limitation.

- Asset – It increases the likelihood that assessments are sufficiently comparable to be treated as alternate-form exams, allowing them to be analysed together for reliability and student ability estimation
- Limitation – Teachers have little latitude to design student tasks. While they can contextualise assessments for local settings, the overall structure and form are tightly constrained

Teachers are also prohibited from re-using items or exams from previous cycles. Although sample IAs and a bank of previously endorsed assessments are available, any submission that reproduces or closely copies earlier materials will not be endorsed. These narrow boundaries leave teachers with limited latitude to design new items that meet specifications without duplicating existing ones. Despite the strong assumption of similarity across IAs, it is unlikely that students gain a significant advantage from access to past materials. Research suggests that pre-examination disclosure of assessment content has little impact on performance (Powers, 2005).

It is unlikely that educators are gaining significant generalisable skill in developing quality assessments given the tight restrictions on acceptable content and format for IAs. They are, however, required to create new IAs each year. Some large-scale assessment programs, such as the Graduate Record Examination (GRE) in the United States, release the entire pool (ETS, 2023) of their analytical writing prompts. The rationale includes reducing candidate anxiety, providing a more accurate understanding of the format and scoring, and supporting timing. Like the IAs, the GRE has a tightly enforced structure for these tasks. GRE test developers also believe that the current pool comprises most of the 'good' prompts that can be written in the specified format. A similar constraint may appear in the QCE IAs over time.

There was no strong rationale provided for the restriction on re-use of previously endorsed items or IA test forms. QCAA now has access to several years of endorsed IAs; these samples could be offered to educators as a library, a tool they can elect to use or not to create their IA. Once built, QCAA could offer educators the option of using a pre-endorsed item or instrument from the library. This would reduce the time educators must invest every year in creating a new IA for Endorsement. A potential technology-enabled system is depicted in Figure 5.6 below.

Such a system uses a combination of GenAI and humans-in-the-loop to classify submitted IAs. Some, potentially a large number, could be automatically endorsed, leaving the Lead Endorsers to utilise their expertise only where needed. This could reduce the burden on educators and casual endorser staff alike, saving money currently spent on casual endorser training, materials and salaries.

Figure 5.6 Technology-enabled Endorsement procedure

Confirmation

Confirmation in the QCE system provides assurance of marking quality and consistency with the specific ISMG. The alignment of the IA with the syllabus and ISMG has previously been verified in Endorsement. Confirmation is not a re-marking process; it is intended to increase the comparability of marked student results on the locally developed assessments. Like Endorsement, there are technology tools that could be used to streamline the process while maintaining its intent. If these tools are acceptably accurate, Confirmation could be extended from a sample of results to the entire cohort with minimal additional investment and time, further increasing confidence in the accuracy and consistency of IA marking.

There are existing content alignment systems that use GenAI to evaluate how accurately submitted content aligns with the framework or syllabus on which the assessment items are based. These systems can provide not only alignment results but also an explanation of the rationale for the classification. Such a system could be used in Confirmation, as shown in Figure 5.7 below.

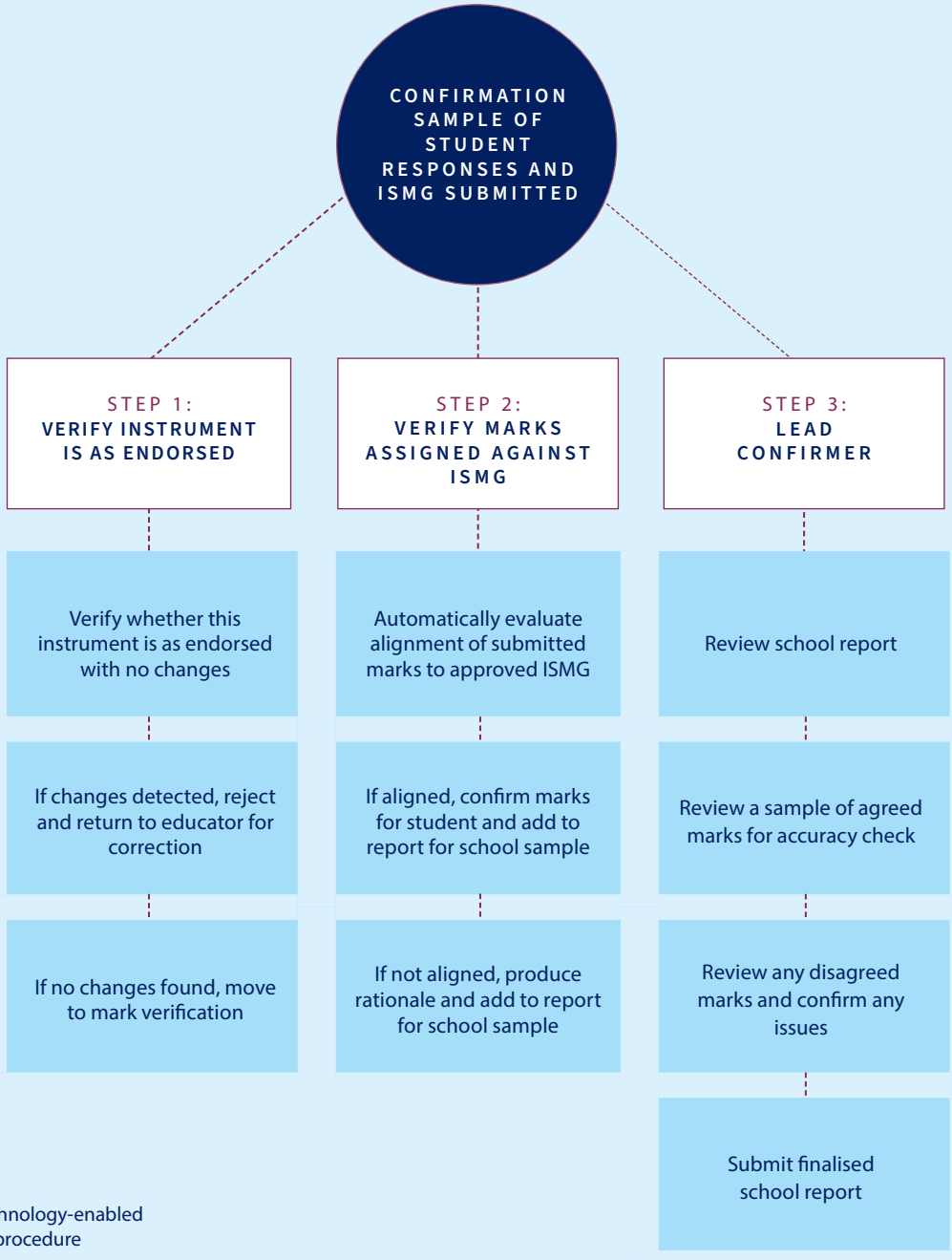


Figure 5.7 Technology-enabled Confirmation procedure

As with Endorsement, the system uses a combination of GenAI and humans-in-the-loop to classify submitted marks as aligned to the ISMG or not. Once trained on a set of expert confirmer decisions, this system could increase the accuracy of Confirmation as its judgements would likely be more consistent. The workload of the Lead Confirmers is reduced, although their expert judgement remains the ultimate decision point. This could ease the burden on educators and casual confirmer staff, using fewer resources with high efficiency to complete the QCE Confirmation tasks.

External Assessment

Two stages could potentially be enhanced by technology in the EA: development and marking. Neither, however, is likely to benefit unless the GenAI system has been specifically developed for the purpose. General use chatbots like Claude or ChatGPT are unlikely to meet the QCAA quality standards and could reduce public confidence, given their tendency to errors and hallucinations. In External Assessment, as in all QCE systems, human experts should remain central to the process.

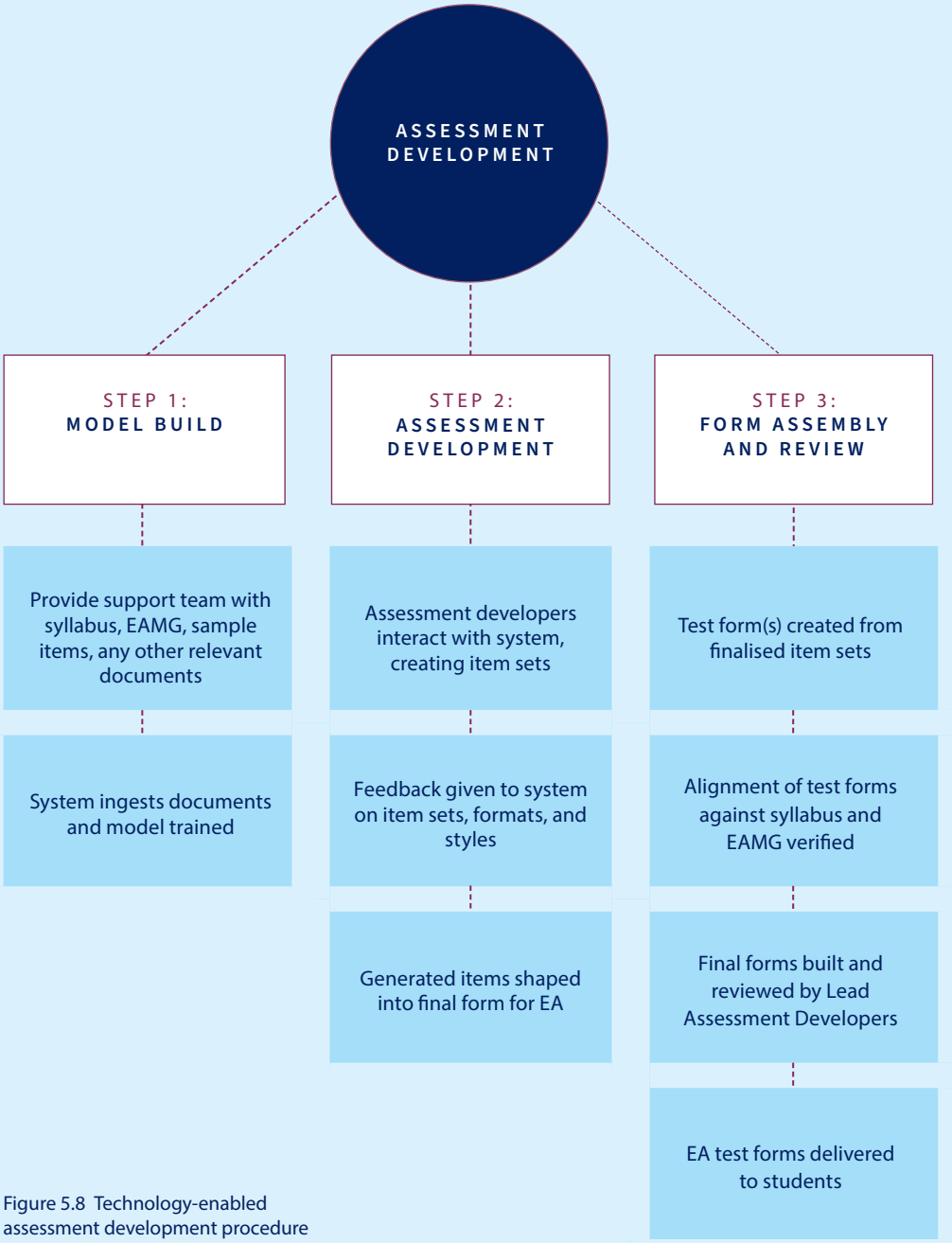


Figure 5.8 Technology-enabled assessment development procedure

Assessment development *(see previous page)*

For assessment development, GenAI tools are now available to support assessment writers. These tools can create items in almost any format: multiple choice, short response, extended essay and complex task assignments. Developers can select the topics and upload the content framework, performance standards, syllabus, marking guides, sample items and any other related documentation to improve system performance before item writing begins. Items can be written to target specific reading levels and cognitive complexity. The 'creativity' setting can be adjusted to keep new items close to existing ones or to allow more differentiated possibilities. As developers interact with the system, it learns their preferences and improves through feedback.

The content alignment system described in the Confirmation section can also be used to evaluate how well the proposed item pool or test form covers the syllabus or framework, supporting appropriate content span and coverage. Once quality assurance reviews are completed, the final EA test forms can be printed or delivered online to students. This process is expected to reduce the time and resources needed to produce EA forms without a loss of quality or fidelity to the syllabus. This process is shown in Figure 5.8

Marking

There are a range of automated marking systems, both GenAI-enhanced and more traditional. These systems have limits in terms of what they can assess effectively. Currently they are not well suited to marking complex performances in areas such as Art, Drama or Physical Education. In most traditional content areas, however, they are highly effective and accurate. Accuracy is improving rapidly as models become more sophisticated and better able to handle complex submissions and rubrics. Advances in optical character recognition now allow handwritten submissions to be digitised and automatically marked without manual key entry, saving significant time and resources. The process is shown in Figure 5.9 overleaf.

As human marking is often the factor that determines the timing of results release, incorporating technology into marking could reduce the time between assessment and reporting of results.

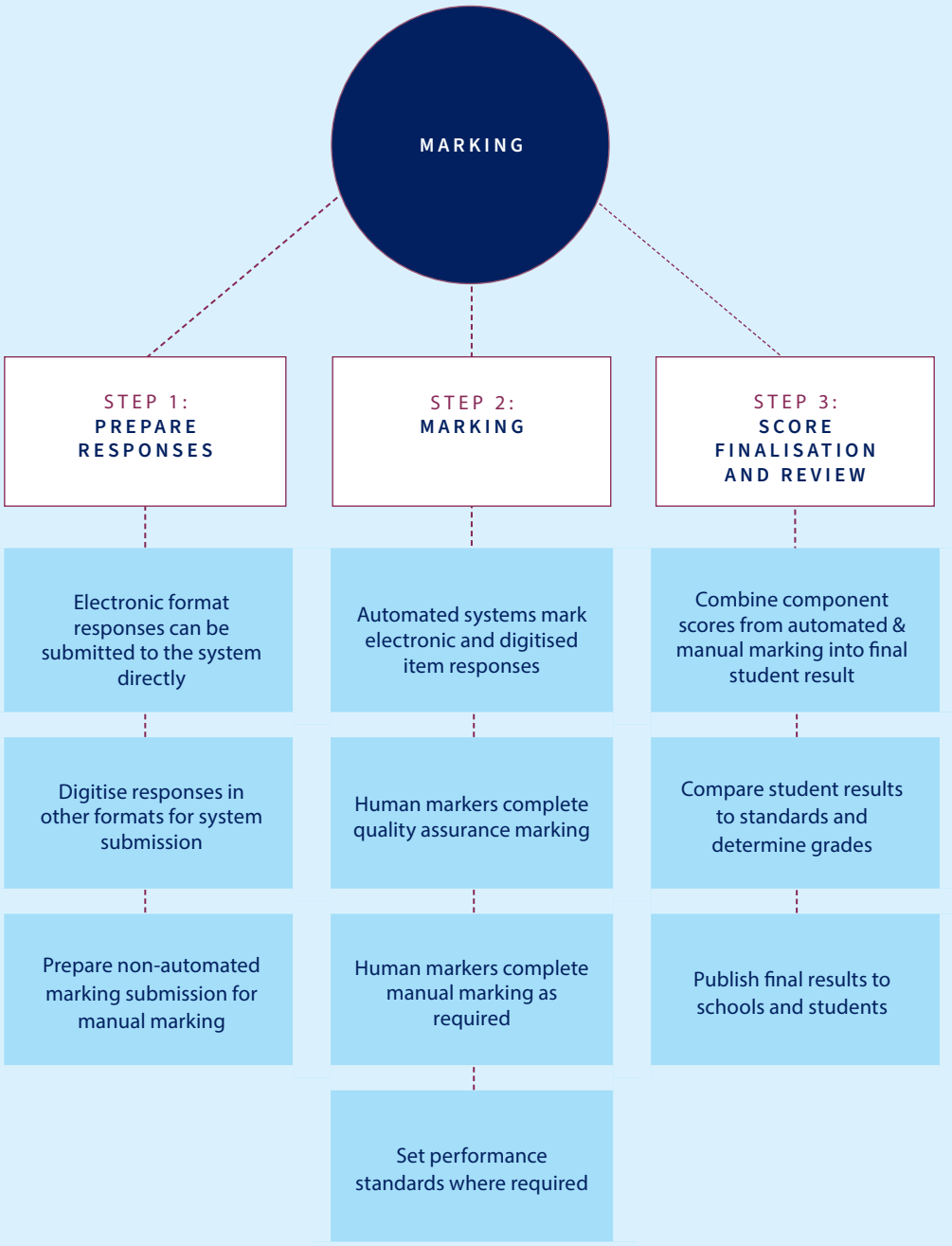


Figure 5.9 Technology-enabled marking

Implementation of advanced technology will require an initial investment of funding. However, there are likely to be significant savings realised from the reduction of casual staff costs, among other factors.

Recommendations

- QCAA to undertake review of incorporating advanced technology into the QCE assessment processes, including considering the value for money/return on investment and complexity of change management associated with the implementation of such technology

Suggested Operational Adjustments

- QCAA to consider improving efficiency of assessment procedures through addition of pre-approved components for use by assessors
- QCAA to consider implementing a process parallel to 'control scripts' for Endorsement and Confirmation, where some expert-scored cases are included, blind to the assessor, in the queue to quantify accuracy and consistency of decisions

Area for review 4.2

QCAA's strategy for the recruitment and retention of casual assessors to ensure there is an available and dependable workforce for each assessment process and activity.

Casual assessors: Summary of evidence

In this area for review we consider casual positions of endorsers, confirmers and markers. Invigilators, who are not assessor staff, were also surveyed. For descriptions of each of these roles, see Area for Review 4.1. More detailed information about the survey questions and responses is provided in Appendix E.

The new QCE system relies on casual assessors for assessment processes and activities. There are no reports of shortages of casual assessors. All processes have proceeded as expected and recruitment strategies have been successful in finding trained staff to complete the tasks. The accessible professional development training modules on the QCAA portal include targeted content for endorsers, confirmers, EA writers and EA markers. Successful completion of these modules and a passing mark on the associated examination is required to apply for the roles. These modules attract fewer respondents than the more general modules on assessment literacy but provide enough staff for each procedure. Successful applicants are required to attend training specifically targeting the precepts, activities and specialised software used for the role. Comprehensive training is led by expert staff, with practical exercises designed to build applied skills in Endorsement, Confirmation, EA writing or marking. New assessors must complete several examples designed with high fidelity to their role. They begin with demonstration, moving through examples as a group, to completing examples on their own, with discussion of the results. Next, they familiarise themselves with features of the QCAA application. Surveys completed by casual assessor staff showed strong intertest in working additional cycles, demonstrating that there is no reason to anticipate staff shortages in the near future.

Casual assessors and other stakeholders in the QCE system expressed concerns about the sustainability of the casual assessor workforce. QCAA staff reported mixed views on which processes could sustainably support these roles over time. Overall, risks to sustainability exist given teacher workloads and the limited number of teachers in some subject areas. Current assessors, however, generally report positive experiences and view the system as sustainable. QCAA should consider further investigation into the sustainability of the workforce to identify potential risks, such as ongoing teacher shortages, particularly in small-enrolment subjects or in regional and rural schools with fewer staff. In these contexts, the pool of potential assessors is very small. Some teachers may be unable or unwilling to participate, creating challenges for scheduled completion. In addition, teachers in small subject areas are likely to know each other, which may introduce biases to the process.

Findings

- QCAA provides accessible training to support educators interested in applying for casual assessor roles in the QCAA portal
- The specialised training for assessors who have completed the online modules and passed the assessment is conducted by expert staff, and is thorough and engaging
- Assessor training effectively develops applied skills through numerous embedded exercises
- Those in casual assessor positions feel that they are well prepared for their roles and that they had adequate time and resources to complete their duties
- Stakeholders in the QCE system expressed concerns about the long-term viability of the casual assessor workforce

Recommendations

- QCAA to collect and analyse data on casual assessor recruitment and retention, with attention to trends over time that may indicate potential threats to sustainability
- QCAA to consider modifying casual workforce schedules to optimise flexibility and improve efficiencies

5.9 Effective documentation

Area for review 5.2

The documentation of processes and procedures to ensure effective functioning of the system and accountability to its stakeholders, including the QCAA staff and school personnel who play essential support roles (e.g., principals' delegates, External Assessment coordinators, assessors).

Governing complex education organisations demands documenting processes and procedures which can support stakeholders' leadership and implementation of reforms (OECD, 2016). Policies further outline which stakeholders are accountable for different processes and outcomes. A healthy organisation will establish guiding principles for documentation of processes, with focus upon clarity, accessibility, trust and accountability.

Policies will usually include audit trail requirements for critical processes, reporting relationships and oversight of responsibilities, as well as documentation of consequences for procedural non-compliance. QCAA has a large and varied library of documentation, encompassing roles and responsibilities. These documents are constructed in a similar format, making locating important information easy and the structure familiar to users. The handbooks and other procedural documents reviewed are comprehensive, detailed and transparent, with diagrams of processes and complex structures to enhance their utility and clarity. The documents reviewed online outline expectations to schools and clarify the roles between different stakeholders. The QCE & QCIA Handbook (2025) outlines quality assurance processes, including defining the partnership between schools and the QCAA: "The quality management system describes the principles and processes required for quality assurance related to the QCIA principles for quality assurance" (QCE & QCIA Handbook, 2025).

The evaluation did not collect quantitative evidence regarding the understanding of users of the system documentation, whether expected procedural actions are taken based on the documentation, or instances in which misunderstandings or errors have occurred been based on the existing documentation. As a result, we cannot comment effectively on the sufficiency or effectiveness of these documents.

Findings

- QCAA provides extensive documentation of all operational procedures and roles in the QCE processes
- The structure and content of the QCAA documentation are presented in a consistent and accessible format

Recommendations

- QCAA to consider reviewing its documentation suite to ensure clear content, accurate alignment with enacted operational procedures and sufficient detail to support implementation

Area for review 5.3

The redevelopment of QCAA's ICT systems to support the new QCE system. This should include an assessment of the reliability, accessibility and usability of the new systems, processes for evaluation or ongoing review of applications, and compliance with Queensland Government policies and guidelines.

The evaluation of complex information technology systems is a highly specialised skill set and is not well suited to a general evaluative study. QCAA has several custom applications in the QCE system which should be evaluated for user experience and effective software architecture and design by a team qualified in these disciplines.[2]

Recommendations

- QCAA to commission a study of the user experience, code architecture, accessibility and functionality of the ICT systems supporting the QCE system from a suitably qualified technology team

[2] QCAA and UoM agreed that QCAA's ICT systems could not be substantially addressed in this engagement.

Area for review 5.4

The suitability of access arrangements and reasonable adjustments to meet the needs of students whose access to education is affected by disability, impairment, medical condition or other circumstances.

As part of the QCE system, QCAA recognises some students will be affected by disability, impairment, medical condition or other circumstances. The Access Arrangements and Reasonable Adjustments (AARA) process is designed to enable these students to participate in assessment on the same basis as other students and to have opportunities to demonstrate their knowledge and skills.

AARAs can be granted to students for both Internal and External Assessments in the QCE process. IAs are managed by schools. They include assessment tasks (Internal Assessment, e.g., a mathematics problem-solving activity) or a timed examination facilitated by the school (internal examination, e.g., an in-class mathematics test). External Assessments are managed by QCAA and focus on timed, invigilated examinations undertaken by all students completing a QCE subject. Different circumstances (e.g., Internal vs External Assessment; Units 1–2 versus Units 3–4; type of required adjustments) call for different types of AARA. Some AARAs can be approved by school principals whereas others require QCAA approval. For example, assessments in Units 1–2 require school-based approvals whereas assessments in Units 3–4 can be either school- or QCAA-based. Students may have several different AARAs across different subjects and assessment tasks. Figure 5.10 demonstrates some examples of AARAs for Internal and External Assessments.

Principal-reported AARA

Internal assessment	Internal and External assessments
• alternative format papers*	• bite-sized food
• assistive technology*	• drink
• computer*	• individual instructions
• assistance	• medication
• reader	• diabetes management
• scribe	• physical equipment and environment
• comparable assessment	• varied seating
• extensions	• vision aides
• variation to venue	

*QCAA approval is required for the CIA in Essential English and Essential Mathematics only

QCAA-approved AARA

Internal assessment – examinations*	External assessment
• extra time	• extra time
• rest breaks	• rest breaks
• alternative format papers*	• alternative format papers
• assistive technology*	• assistance
• computer*	• assistive technology
	• computer
	• reader
	• scribe
	• variation to venue

*QCAA approval is required for the CIA in the Essential English and Essential Mathematics only

Figure 5.10 Examples of principal-reported and QCAA-approved AARA

QCAA provides some guidelines for circumstances that are (e.g., short- and long-term medical conditions, funerals/sorry business, weather-related events) and are not (e.g., family holidays, timetable clashes) eligible for AARA, but also allows for flexibility in the system so that other circumstances can be reviewed. Judgements about access and appropriate adjustments are also made on a case-by-case basis. For example, multiple students with the same disability may require different assessment adjustments; an individual student may need different adjustments for different subjects or assessment types; or an individual student may change in adjustment requirements throughout senior secondary school.

AARA applications are submitted by school staff through the AARA app in the QCAA portal. This system can be used to document principal-reported AARAs (i.e., not requiring QCAA approval) and to submit applications for AARAs requiring QCAA approval. Different types of supporting documentation are required for QCAA-approved AARA applications. QCAA has a dedicated AARA unit to review and manage applications needing QCAA approval, develop resources for schools and teachers, and assist in the AARA application process.

Providing access arrangements and reasonable adjustments for students affected by disabilities, impairment, medical conditions and other circumstances is a necessary component of a high-quality assessment system. In line with the Queensland *Human Rights Act 2019* (Queensland Human Rights Commission (QHRC, 2019)) and the United Nations Convention on the Rights of the Child (United Nations Convention on the Rights of the Child (UNCRC, 1989)), all children, regardless of their abilities and needs, have a right to education. This education should be accessible (UNCRC, 1989), article 28) and appropriate to the child's needs (QHRC, 2019), section 36[1]). Ensuring assessments are accessible and adjustments are implemented to meet individual students' needs and circumstances is an important way to provide appropriate special care and support resources that facilitate active participation in the community ((UNCRC, 1989), article 23).

These adjustments are also important points of compliance with the Australian *Disability Discrimination Act 1992* (Australian Government, 1992) and the *Disability Standards for Education* (Australian Government, 2005). According to the *Disability Standards for Education*, education providers have an obligation to:

- make reasonable adjustments which allow students with disability to participate and learn on the same basis as students without disability
- prevent harassment and victimisation of students with disability
- consult regularly with students with disability and their parents and caregivers to develop reasonable adjustments and address concerns as they arise

It is important that QCAA's AARA processes account for and comply with these standards. QCAA's AARA arrangements are similar to those of other jurisdictions in Australia, including NSW and Victoria. Each system includes eligibility criteria in line with ways to recognise arrangements for in-school examinations (e.g., principal-reported AARAs).

In this evaluation, we review:

- summary information on AARA applications
- the design of AARA policies, process and adjustments, particularly regarding the extent to which information and support for AARA are accessible and readily available
- the fidelity of implementation of AARA processes
- the effectiveness of AARA adjustments

Summary of AARA applications

In 2024, for IAs and internal examinations:

- 6,800 students had principal-reported AARAs (extensions) and/or comparable assessment tasks delivered for their IA program
- 6,153 students had principal-reported adjustments for IAs that were completed under timed conditions (i.e., internal examinations)
- 5,652 students had QCAA-approved adjustments for IAs that were completed under timed conditions (i.e., internal examinations)

These student groups may partially overlap, as students may have both principal-reported and QCAA-approved AARAs in their IAs and examinations.

Data suggests a slight increase in the number of QCAA-approved AARAs for Internal Assessments from 2021 to 2024 (~2%). For principal-reported AARAs, rates in 2024 were similar to 2023 (12% versus 13%) but lower than 2022 (17%). It is possible that 2022 saw a temporary increase in AARAs for IAs due to the COVID-19 pandemic. Future work may consider examining the extent to which these trends hold over time.

For External Assessments, in 2024 there were:

- 4,115 students with a principal-reported AARA
- 4,607 students with a QCAA-approved AARA

Both principal-reported and QCAA-approved AARAs for EAs demonstrate slight increases (~2%) since 2021. Again, these student groups may partially overlap, as students may have both principal-reported and QCAA-approved AARAs in their External Assessments. These groups would also overlap with the Internal Assessment AARA rates reported above.

In 2024 there were 12,682 students engaging with the AARA system overall, with 6,505 students with QCAA-approved AARA. The acceptance rate for QCAA-approved AARAs is high, with 94.1% of 2024 applications having at least one adjustment approved. This acceptance rate has been relatively consistent since 2021 (93.8%). On the occasions where QCAA-approved AARA applications are not approved, the application is sent back to the school. It may be, for example, that the application needs further information that the school can provide. When schools resubmit an AARA application, this counts as a 'new' application in the reporting system. Thus, when accounting for applications that are rejected but then resubmitted and approved with

further information or justification, acceptance rates may be slightly higher than the 94.1%.

There were a variety of conditions across AARA applications (e.g., attention deficit hyperactivity disorder, hearing impairments, autism spectrum disorder). The cognitive and social/emotional categories account for the largest proportions of QCAA-approved AARA applications for EAs. The most common approved adjustments from the QCAA-approved applications were extra time, rest breaks, assistive technology and assistance.

Although the available data does provide relevant information regarding principal-reported and QCAA-approved AARAs, QCAA may consider updating reporting standards to add more consistency and transparency to the number of submissions, approvals and students engaged in the AARA process. There are differences in terminology and reporting across annual reports and documents that make it difficult to understand trends across years. Greater consistency in reporting standards would help with tracking changes and challenges in the AARA process over time.

Design of AARA policies, process and adjustments

The QCAA website provides information about the AARA policy, required documentation and application process, supported by downloadable fact sheets. Phone and email contact details for the AARA team within QCAA are prominently displayed on the website and information directs readers to this team for queries. The information available is clear and comprehensive.

In the school case studies, most participants who had experience with the QCAA AARA unit found the team helpful:

> *They [QCAA AARA team] provide exceptional support for us putting in the applications, they provide amazing support during that external exam time; like on the phone, making decisions, it's next level.* (Head of Department)

Nearly all QCAA staff who support schools in the AARA process reported that teachers found the AARA resources useful. Overall, it appears that the information available is comprehensive and easily accessible online, and support is available if needed.

The design of AARA policies and processes is aligned with the Australian Nationally Consistent Collection of Data on School Students with Disability (NCCD; Department of Education, Skills and Employment, 2013) policy and is in keeping with those offered in NSW and Victoria. Some notable examples of alignment include that imputed disabilities are acceptable for AARA applications (QCAA, 2023g) and that GP certificates are sufficient evidence for pre-existing conditions (QCAA, 2023d).

The process is also inclusive of some cultural protocols (e.g., sorry business and sad news – (QCAA, 2023i)) and illness/misadventure (QCAA, 2023f). This provides additional flexibility in the AARA process that is beneficial for students (Cumming et al., 2013; Page et al., 2024).

Although the application process is similar to the NCCD, stakeholders reported mixed experiences when it came to engaging with the application process. Many reported that the QCAA portal and AARA app for managing AARA applications was suitable. However, experiences of school staff, parents and students of the AARA process were mixed.

Some school staff members, students and parents reported that the process was smooth to implement, streamlined, and was "brilliant" at supporting students. Many general survey respondents wrote "AARA" as their response, or part of their response, when asked what is working well with the new QCE system. Multiple students who had been granted adjustments spoke about the relief they felt knowing that the adjustments were available, and principals noted the flexibility of the system.

> *I found it was a relatively smooth process. All I needed was to talk to another GP and him to write me a report. So, I feel like it was a relatively fair and simple process.* (Current student)

> *I think the AARA itself is fine, somebody gets sick, somebody has an ongoing medical condition, somebody has sorry business in their family, I think that's all really good.* (Principal, remote school)

At the same time, other school staff members, students and parents reported difficulties with the AARA process and with accessing resources. Many school staff members and some parents and students wrote "AARA" as their response, or part of their response, when asked what is *not* working well with the new QCE system. Written feedback suggested that difficulties surrounded meeting AARA requirements, completing core tasks, and accessing resources:

> *Despite AARA adjustments those with common disabilities are doomed when they avoid high scaling subjects because they don't exam well.* (School staff member)

> *Core completion can also be an issue for students. Illness and Misadventure AARA applications are very human resources heavy and take time to complete and issue extension dates. This elongates a process and the setting of a new revised due date. Catering for the diversity of Adjustments in formal exams is also challenging with space and technology. The provision of appropriate spaces to hold exams is also limited and takes away the use of physical resources such as the Hall for schools for an extended period of time.* (School staff member)

Some reported difficulties reflect misunderstandings about the AARA process and its documentation requirements. For example, issues were sometimes conflated with other aspects of senior secondary schooling, such as the difference between tertiary entrance and the QCE (i.e., high-scaling subjects) or between school-related issues and the AARA process (i.e., school time-tabling and physical space). There was also confusion about when specialist documentation was required, the level of documentation needed for students with permanent disabilities or involved in sorry business, and the use of imputed disability.

> *In grade seven, … I was diagnosed with dyslexia. … But since it [the diagnosis] was outdated, … and since there's no dyslexia testing places in [rural town], I would have to go to Brisbane and get on a waitlist and wait to get another testing, and it was just impossible. So, I ended up just giving up.* (Current student)

> *I know students who are well and truly diagnosed as being on the autistic spectrum, ASD, for them to go and be diagnosed again ... It makes the experience really lengthy, it's a burden on parents, it's incredibly expensive.* (Head of Department)

> *We have a lot of Indigenous [students] who when they go out on sorry business, they can be gone for up to a month. And sometimes we don't even know when they're returning. So there's no [AARA] at play for an Indigenous student.* (Remote school leader)

> *I can impute someone with a substantial disability under the NCCD model. But not under the QCAA [AARA requirements].* (Guidance officer, low SES school, rural area)

Although all of these were explained in the AARA documentation reviewed above, the information is not reaching some participants.

The QCAA staff survey further reflected these mixed experiences. Several QCAA staff who actively supported schools through the AARA process were asked whether a) schools implemented AARA processes in accordance with the documented processes, and b) schools were confident in their ability to do so. Results showed that only approximately half of these staff felt that schools implemented the process with fidelity and confidence. Although the sample is small, these staff had contact with several schools. The alignment of these mixed views with those of AARA end users suggests that the process may be more difficult to navigate for some schools than others.

Misunderstandings may also vary by school context. Existing QCAA data suggests that government schools were under-represented in QCAA-approved AARAs. Based on the proportion of applications in 2024, about 3,593 students from government schools were expected to have QCAA-approved AARA. Observed rates show only 2,537, which is 29.4% lower than expected. In contrast, Catholic schools (expected = 1,328; observed =1,821; 13.5% higher) and independent schools (expected = 1,585; observed = 2,147; 14.1% higher) exceeded expectations. This pattern was consistent from 2021 to 2024.

There could be several reasons for this discrepancy. Government schools may be less likely to submit applications due to less familiarity with the process or fewer resources to prepare submissions. Analysis of the 2023 general subject enrolments by QCAA district also revealed significant differences between schools in Brisbane Central and those in regional, rural and remote areas. It may be that regional, rural and remote schools, as well as government schools, have less awareness of AARA resources.

The extent to which these difficulties are systematic is an important area for future review. QCAA should investigate whether differences in AARA experiences are linked to factors such as reasons for AARA (e.g., sorry business, imputed disability, conditions that require specialist diagnosis), school location (e.g., rural, remote schools) or school type (e.g., government schools).

The adjustments provided through AARA appear suitable. Psychometric analyses indicate no statistically significant bias between the performance of students who receive assessment adjustments and those who did not. This provides validity evidence supporting comparable inferences from both groups.

A small proportion of staff and parents requested that QCAA consider untimed assessment as a future option to better support students with high anxiety. Current data suggest that students with extra-time adjustments often do not use the extra time (QCAA, 2024b). In 2024, of the 4,020 students approved for extra time in EAs only about 6% used all of it and 30% did not use any. Among the 2,158 students approved for rest breaks, fewer than ten (0.4%) used all their time and 55.1% did not use any. Researchers have noted that the perception of extra time may itself be valuable, helping students to ease anxiety and improve motivation (Elliott & Marquart, 2004; Wei & Zhang, 2024). As one staff member summarised:

> *It's [AARA's] really positive. And we do find, for a lot of students – so they might identify rest breaks or extra time. They don't actually take it, but the fact that they've got it there means that they're calm enough to be able to perform well in their assessment.* (Head of Department)

Determining whether the adjustments provided to students are reasonable for their circumstances is beyond the scope of this evaluation. In future reviews, QCAA could consider how well AARA processes and adjustments align with the *Disability Standards for Education* (Australian Government (2005); see also Cumming et al. (2013); Page et al. (2024). Particular attention could be given to the extent to which students and their associates are consulted in determining appropriate adjustments, as research has noted that communication between schools, parents, and students is a crucial component of the adjustment process (Page, et al., 2024).

Findings

- AARA information available is comprehensive and easily accessible online, and support is available if needed
- End users of the AARA process and procedures have mixed experiences
- No statistically significant bias in results was found between students who receive assessment adjustments and those who do not

Recommendations

- QCAA to consider updating reporting conventions to add more transparency to the number of submissions, approvals and students engaged in the AARA process
- QCAA to investigate the extent to which experiences with the AARA process vary by different student or school factors (e.g., reasons for AARA, school location or school type)

Area for review 5.5

Whether the processes and systems for calculating and reporting subject results are well-defined, efficient and transparent.

The QCAA has established clearly documented processes for calculating and reporting subject results in its official guidelines online (QCAA, 2020). The guidelines state how subject results are calculated for students completing studies in general, general (extension), and general (senior external examination) subject syllabuses in the QCE system. Publishing this level of detail increases clarity and transparency in what can otherwise be a complex process for educators and the public.

The QCAA's guideline outlines a systematic approach for calculating general subject results using the Rasch model, a highly respected method for analysing educational assessment data worldwide. This model creates a logit scale on which both student abilities and assessment item difficulties are placed, allowing fair comparison across different assessments. The process involves three key analyses: first establishing the scale using IA results, then placing EA results on this scale and, finally, creating a consistent year-to-year comparable EA score. This 'divide and conquer' approach separates key stages of the calculation so that each can be completed optimally. It reduces the likelihood of errors that might occur if all analyses were conducted simultaneously and allows direct addition of Internal and External Assessment marks to calculate final subject results while maintaining comparability between years despite variations in EA difficulty.

For general subjects, the calculation process involves the IAs (developed by schools but endorsed by QCAA) and the EA, with results combined to calculate a final subject result. Schools make judgements by matching evidence in student responses with syllabus standards, following guidelines for both Unit 1–2 and Unit 3–4 assessments. The process defines how to handle special cases, such as non-submissions of student responses, which schools record as 'not rated' results.

The guidelines also detail how reporting standards (A–E) are established through an item-descriptor matching method, where subject experts map assessment items to syllabus standards, allowing cutoffs to be determined on the logit scale and then translated to raw scores. Special procedures address situations including missing results, senior external examinations, subjects with multiple EAs, alternative sequence syllabuses, subjects with different IAs but common EAs, and illness and misadventure scenarios. A Ratification Committee, comprising representatives from the QCAA Board, schooling sectors and independent experts, reviews and endorses these processes each year, making variations as necessary to ensure fairness while maintaining the integrity of the system.

An evaluation of the QCAA documents and guidelines shows evidence of transparency in the system for students as well. Students can access their subject results through their online learning accounts. They can also apply for verification of their statement of results or request reassessment or senior external examination scripts if they believe an error has occurred (QCE & QCIA Handbook, 2025).

QCAA also produces annual subject reports for schools that provide an overview of key outcomes in student achievements. These reports support the teaching, learning and assessment and serve as an additional signpost of a systemic approach that fosters organisational learning. Unlike other jurisdictions globally, Queensland does not publish individual student results or rank schools. However, QCAA does publish qualitative and quantitative reports on assessment and assessment outcomes for the public.

Overall, QCAA documents clearly outline the processes and systems for calculating and reporting subject results, and the reporting approach facilitates clear communication with students and their guardians.

Findings

- The processes and systems for calculating and reporting subject results are well-defined and transparent
- Publicly available QCAA guidelines detail how grade reporting standards are set
- The approach for calculating general subject results uses the Rasch item response theory (IRT) model, a well-regarded, technically sound method for analysing educational assessment data
- Use of the Rasch model creates comparable scores both between years and across Internal and External Assessment marks used to calculate subject results
- A staged approach reduces the complexity of concurrent analysis and the likelihood of error, improving efficiency and accuracy

06

Monitoring, review and system impact

6. Monitoring, review and system impact

The QCE is intended to deliver a "world-class curriculum", where "teachers have a strong conceptual and operational understanding of senior curriculum and assessment processes and improved capacity to design and deliver high-quality assessment" (QCAA, 2017) The extent to which this has been achieved is reviewed for curriculum and syllabus provision in Chapter 4 and for assessment in Chapter 5 and Appendix B. In both cases, the initial implementation of the new QCE has been followed up with review and changes in policy and practice to enhance the system. This chapter reviews the way in which the overall system is monitored to identify opportunities for improvement and the way in which data and other information are used to support the process.

It presents findings regarding these elements of the new QCE system identified in the terms of reference:

- governance (Area for review 6.1)
- review and improvement (Area for review 6.2)
- availability and accuracy of data (Area for review 6.3)
- level of public confidence in system (Areas for review 7.2)
- academic integrity and responsible and ethical action (Area for review 7.3)
- the result of increased advice to teachers (Area for review 7.4)
- impact of new accreditation and training on teachers' practice (Area for review 7.5)
- fairness and accuracy of subject results (Area for review 7.8)

Other areas for review have been excluded by agreement between QCAA and The University of Melbourne team on the grounds that there were insufficient data collected to support valid inferences about them:

- stakeholder understanding of the system (Area for review 7.1)
- impact of reduced summative assessments (Area for review 7.6)
- flexibility to cater for individual student's needs (Area for review 7.7)

6.1 Governance in evolving systems

Area for review 6.1

Ongoing governance arrangements to ensure the system is monitored, reviewed and continues to evolve to meet changes in the wider educational environment.

6.1.1 Governance in evolving systems

Education governance involves structuring how education activities are organised and by whom they are performed (Bardhan, 2002). Over recent decades there has been an expectation that the role of governments in education would diminish as new governance arrangements emerged expanding the involvement of external and non-state actors alongside traditional schools. However governments remain responsible for engaging partners and bring their conflicting interests into a coherent system (Capano, 2011). Making governance arrangements clear to stakeholders translates into better education outcomes (King & Özler, 2005).

Previous studies on governance of complex education systems highlight the importance of balancing top-down guidance and control with local autonomy creating supportive networks building capacity and developing sustainable successful implementation strategies (Burns et al., 2016; Hopfenbeck et al., 2015; OECD, 2018). Clear communication and trust between governance levels are essential for successful implementation as is long-term planning for sustainability.

International comparative studies of student achievements such as TIMSS, PIRLS and PISA (Hopfenbeck & Görgen, 2017), and particularly their country rankings (Grek, 2015; Hamilton et al., 2015), have pushed systems away from strict process regulation towards goal setting by governing bodies, with decentralisation of responsibility for pedagogy and, to some extent, curriculum content (Soguel & Jaccard, 2008). This approach has been described as "competition and choice", "standardisation", "borrowing of change models from the corporate world" and "test-based accountability policies" (Sahlberg, 2016, pp. 133–136).

Governments worldwide have also expanded access to education (Saguin, 2019), often dramatically increasing enrolment and bringing into the system formerly neglected students, many of whom may have specialised needs. This expansion has strained resources in already high-demand systems and heightened tension between egalitarian goals – seeking equality of opportunities for all and equity of outcomes – and meritocratic goals that emphasise enhancing individual achievements (Soguel & Jaccard, 2008).

State and national curriculum and assessment agencies often have in-house expertise but also draw on external specialists. Australia's national agency, ACARA, develops and revises national curriculum by having its staff engage with state and territory jurisdictions and higher education experts. Its work on assessment is supported by consultation through its National Assessment, Data, Analysis and Reporting (NADAR) Reference Group and technical advice from its Measurement Advisory Group (MAG), which includes expert psychometricians in Australia and overseas. All of ACARA's work is supported by its First Nations Australians Advisory Group and its Students with Disability Advisory Group (ACARA, 2025).

6.1.2 Evaluation of governance of the QCE system

Implementation of the new QCE was overseen by a Senior Secondary Assessment Taskforce established and chaired by the Minister for Education. It included representatives from the schooling sectors, parents, teachers, principals, unions, the tertiary sector, the QCAA and the QTAC (QCAA, 2025g). It established 22 subject working groups to provide expert advice on curriculum and assessment. The groups conducted a survey of teachers, heads of departments, and principals and deputy principals that showed that "there was overwhelming support for preserving and strengthening school-based assessment and the introduction of complementary External Assessments but some variation in views about the weightings to be applied to the two components". These results were published on the QCAA website.

QCAA provides top-down prescription on curriculum through its syllabuses and control over external examinations and key aspects of school-based assessment. This is balanced by a level of local autonomy supported by professional development activities to build local capacity. QCAA also plays a critical gate-keeping role by certifying students' records of studies and achievements through the QCE and QCIA certificates. In these ways QCAA supports successful implementation in the manner commended in the research evidence (Burns et al., 2016; Hopfenbeck et al., 2015; OECD, 2018).

The ongoing operation and development of the QCE system is managed by QCAA, which is governed by its own board. QCAA maintains structured relationships with its key stakeholders, provided at a high level by the board. The legislation that established QCAA provides for the following seven board members:

1) the Chief Executive of the Department of Education, or a nominee of the Chief Executive

2) the following persons appointed as members by the Governor in Council:

    a) the Executive Director, or nominee, of the Independent Schools Queensland (ISQ)

    b) the Executive Director, or nominee, of the Queensland Catholic Educational Commission (QCEC)

    c) one nominee of the Minister having relevant corporate, governance or financial qualifications or experience

    d) three other nominees of the Minister having the qualifications, experience or standing the Minister considers relevant to the functions of the authority

Two of the three members in the final category nominated by the Minister are currently senior officers of the Queensland Teachers' Union and the Independent Education Union, giving teachers a voice in the governance of QCAA. There is a dilemma when creating boards in balancing specialised understanding of the system and independent external perspectives. The seven-member QCAA board replaced a 20-member board of the former Queensland Studies Authority. The reduced board is, no doubt, more manageable but it has very little external perspective.

Other agencies use similar structures. For example, the Office of Qualifications and Examinations Regulation (Ofqual), which regulates qualifications, examinations and assessments in England, has external experts on its board, including an Australian psychometrician with substantial US experience. Ofqual also runs an extensive research program supported by its Research Advisory Board (Ofqual, 2025). The governance structure with the QCAA Board and its Standing Committees works well but could be enhanced by further external perspectives, including on measurement issues and on the development of a research agenda.

VCAA has an eight-member board, three of whom are senior managers of school systems. The others cover public administration, higher education and business (VCAA, 2025). The NSW Education Standards Authority (NESA) has a 14-member board with five from education systems, five from schools or with a strong school background and four from higher education. Most of the board members in the VCAA, NESA and QCAA have diverse backgrounds, so it is difficult to place all of them in the four categories – system, school, higher education and business – but all are, not surprisingly, dominated by school-level interests.

The QCAA Board has four standing committees established to support consultation processes and ensure quality of services in areas signified by their titles:

1. Audit and Risk Committee
2. K-10 Curriculum and Assessment Committee
3. Senior Secondary Curriculum, Assessment and Certification Committee
4. Ratification Committee

The K-10 Curriculum and Assessment Committee is outside the scope of this evaluation.

The Audit and Risk Committee's role is to provide independent assurance to the board in the area of risk management. It is a three-member committee with one external member with governance and financial expertise. Sample agendas reveal the committee's attention to internal and external risks and to some specific administrative issues. QCAA's Strategic Plan (QCAA, 2025g) identifies two newer risks related to rapidly shifting technology: data security and the transformation of information technology. The latter encompasses complex issues such as the ethical use of AI and Large Language Models (LLMs) by students, teachers and systems.

There are also risks associated with workforce demands, including upskilling existing staff and the recruitment and management of casual staff required by QCAA. This latter risk was reflected in comments across stakeholder groups, with QCAA staff expressing uncertainty about the long-term sustainability of the current approach and concerns about the timing and the demand load on the casual staff for Endorsement, Confirmation and External Assessment marking.

School leaders recognised that some risks that are beyond QCAA's control. Most notably, they saw teacher shortages as a factor that makes it harder to maintain the new system. This perspective was echoed by external stakeholders, who raised concerns about teachers working outside their area of subject expertise.

It is likely that these risks will converge to some degree, with complex technology and AI-based systems set to alleviate some of the demands on staff skills and time. The high visibility of these risks in the public QCAA strategic plan indicates an intention to monitor and address them as part of the QCE system's evolution.

The Senior Secondary Curriculum, Assessment and Certification Committee has 18 members who all have a strong base in school education, including four senior QCAA staff. It oversees extensive teacher professional development programs, with the associated resources, guidelines and advice, as well as quality assurance supporting high-fidelity implementation of the curriculum and assessment systems. Sample meeting agendas for this committee include reviews of proposed new curricula and draft syllabuses and reports on quality assurance on Internal Assessments. They also include reports on practices in other state curriculum and assessment authorities.

The Ratification Committee has 10 standing members with capacity to include others as required. Three of them are senior QCAA staff, with one member from each of the three schooling sectors, and two independent experts. The committee oversees the processes for the treatment of exceptional circumstances that may impact the calculation of subject scores (e.g., for students with missing assessment results, or where atypical events occurred that may have affected the administration or results). This committee also examines anomalous patterns in results from schools' Internal and External Assessments, supporting QCAA in resolving these issues and working with schools to avoid mismatches in future. Sample meeting agendas for this committee include consideration of a new model for identifying schools with unexpected differences between Internal and External Assessments and reviews of data from schools with unexpected differences in particular subjects.

A clear conclusion on the efficacy of the work of these standing committees cannot be made from the information gathered; however, the sample meeting agendas indicate that committees are focused on content appropriately within their terms of reference. To ensure continued alignment with QCAA goals, ongoing monitoring of their efficiency and effectiveness could be incorporated as a routine responsibility for senior QCAA staff and the board.

Findings

- The governance structure with the QCAA Board and its Standing Committees works well in monitoring and improving the QCE system

Recommendations

- The Queensland Government consider the inclusion of more external members on the QCAA Board, without necessarily increasing the size of the board

6.2 Monitoring and improvement of QCE system

Area for review 6.2

The use of data and other information to monitor the system and make improvements.

QCAA established a commitment to monitoring the implementation of the new QCE. A review of the first cycle for Year 11–12 students was undertaken immediately after the first cohort of students graduated from the new system and was published in August 2021 (QCAA, 2021a). The review reported that schools understood the Endorsement process, teachers' assessment literacy had steadily developed, recruitment and training of assessors had been successful and the resources offered to schools had been appreciated (p. 2).

This early review led to immediate improvements for the next cycle, with publication of revised assessment samples, new app functionality, updated user guides and additional professional development for the Endorsement process, and new supporting resources, revised timelines and new functionality for the Confirmation process. For Internal Assessment, additional student work samples were published to demonstrate the qualities of student work at various performance levels. For External Assessment, new resources for schools and students, new professional development for teachers, streamlined processes for schools, and improved recruitment, training, communication and remuneration arrangements for markers and invigilators were implemented (pp. 2–3).

QCAA developed an explicit plan to implement the changes and introduced a standing item on the QCAA Board meeting agendas to ensure appropriate ongoing monitoring of the change and QCAA's accountability for them. Implementation of the new QCE system involved considerable change in curriculum and Internal Assessment, along with the introduction of External Assessment. A speedy initial review, followed by this more extensive evaluation, has provided QCAA staff and the board with comprehensive information on progress and a foundation for improvement.

Improvement is also supported by QCAA's quality assurance processes for Internal Assessment in applied, applied (essential), general and general (extension) subjects and in short courses. The purpose of these quality assurance processes is to enable schools to develop and administer assessments and ensure consistency of judgements about student responses. Further, these quality assurance processes facilitate the continuous improvement of assessment practices in schools, by functioning as formative processes in feedback loops.

QCAA staff have positive perceptions of the monitoring and evaluation of the QCE system but they also note the complexity created by the number of steps and stakeholders involved, as well as the ongoing need to maintain knowledge and capacity across all participants. These factors are seen as risks for the system's long-term sustainability. Several staff members also mentioned that some data are available to schools but not to QCAA staff, making it difficult for staff to help schools with certain inquiries. This includes, for example, access to information about schools' IA results or enrolment trends. Balancing the needs for school and system-level data can be a challenge because of the need to adhere to privacy legislation.

When the new QCE system was implemented, there was a strong focus on service to schools and consultation with key stakeholders (e.g., school leaders). This is corroborated by evidence gathered from confidential QCAA staff interviews and survey data collected from school leaders. Schools' use of QCAA services, in turn, depends on good, clear communication. An encouraging finding is that a clear majority of school leaders agreed QCAA is responsive to feedback from schools.

Findings

- Implementation of the substantial changes with the new QCE has progressed well with adjustments being implemented where monitoring has already revealed opportunities for improvement

6.3 Availability and accuracy of data

Area for review 6.3

Availability and accuracy of data and other information to support schools, QTAC and other stakeholders, to ensure transparency and public accountability.

6.3.1 Design of data and information processes

A summary of evidence of the use of data to monitor the system externally is provided below. In this instance, we refer to the data teachers receive from the analytics dashboard related to student learning, and information provided to external stakeholders (e.g., subject completions/ enrolments).

The design of QCAA's data and information processes were well received by teachers, schools and external stakeholders. General survey results show that participants found the analytics dashboard easy to access, understand and use (see Figure 9.54 in Appendix E). Qualitative data supported the general survey results, with teachers, principals and heads of departments reporting that the analytical dashboard provided them with better student performance insights that positively impacted teaching and learning. Additionally, data from school leaders indicated that a majority of respondents found the portal easy to use (see Figure 9.55, Appendix E).

QCAA staff reported that there are professional learning opportunities and other means to assist schools in accessing data. Therefore, from the perspective of both the external users of the data (e.g., teachers) and those managing it (e.g., QCAA staff), the logistical design of data and information processes in the new QCE system is straightforward.

6.3.2 Effectiveness of data and information for external stakeholders

The effectiveness of the data and information processes provided by QCAA was also well received. Teachers and school leaders reported that the information provided by QCAA was relevant and informed their teaching practices (see Figure 9.56, Appendix E). Both the analytics dashboard and the subject reports were highlighted as providing key information for teachers to help their students. Teachers reported that these tools helped them focus on areas that students struggle with. Overall, the information provided by QCAA is largely relevant to student learning and teachers' professional practices.

The stakeholder consultations revealed a nuanced perception among teachers and school leaders. They were generally satisfied with the data and information provided. The only comments made was a perceived disparity in access to quality assurance and marking criteria information, noting that teachers involved in Confirmation, Endorsement, and External Assessment marking training appeared to receive more comprehensive professional guidance compared to the broader teaching cohort. While the actual substantive differences in information access may be limited, the perception itself represents an insight into potential communication gaps and suggest an opportunity for strategic improvement in communication protocols.

Other external stakeholders (e.g., QTAC, teachers' union, Australian Skills Quality Authority [ASQA], teacher education institutions) reported that the data they received was of high quality. However, these stakeholders also made several recommendations relevant to their own needs and uses of QCAA data. Although some of this information appears to be publicly available through various reports and resources provided by QCAA (e.g., QCAA annual reports list the number of assessors in Endorsement/Confirmation/marking, annual subject reports provide completion rates for each subject), these external stakeholders may be unfamiliar with where to find this information. For example, the teachers' union recommended releasing annual data on the rates of teachers participating in Confirmation, Endorsement and marking. ASQA recommended upgrading IT systems so that relevant data could be compiled and sent directly rather than through multiple systems and channels, such as student enrolments, assessment outcomes and compliance records, to improve efficiency, accuracy and timely reporting. Finally, teachers' education institutions recommended releasing more data on student subject choices to inform future workforce planning.

Building an understanding of data that are needed by external stakeholders could enable more efficient information exchange. QCAA could consider working with these external stakeholders to identify which data are most useful to them and help them to understand where to find such information.

Findings

- The design of data and information processes in the new QCE system is straightforward with support available when needed
- Information provided to teachers is relevant to teachers' professional practice and students' learning
- Some teachers perceive that those who do not participate in Confirmation, Endorsement and external marking training do not have access to the same information as those who do
- Some external stakeholders suggest that data could be provided more directly and efficiently for their organisational needs

Recommendations

- That QCAA seek to reduce the information gap between those who participate in Confirmation, Endorsement and External Assessment marking and those who do not
- That QCAA work with stakeholder groups to identify which data and information are needed and how to communicate that data efficiently

6.4 Level of public confidence

Area for review 7.2

The level of public confidence in senior assessment and certification since the new system commenced.

Matters and Masters (2014, pp. 37–38) identified threats to public confidence in the previous system – namely, issues with perceptions of fairness, particularly regarding the comparability of scores across different contexts. Although the incorporation of teachers' own judgements about students' learning was seen as a strength of the previous system, the lack of common assessment tasks made it difficult to be certain how to compare students in different contexts. In addition, they noted that the previous system for using results for tertiary selection was overly complex with a lack of transparency in the calculation and use of assessment outcomes to create tertiary entrance scores (p. 58). That is now clearly a matter for QTAC, not QCAA.

The report recommended that students should have assessments in common so that the comparability of students across contexts was clearer. Within each subject, this included assessments that would be administered by teachers and schools internally as well as at least one externally set and marked assessment. The report also recommended simplifying the way student results were calculated under the new QCE system. It was envisioned that addressing these and other ongoing concerns with the previous system (see Chapter 3.1 Rationale and foundation concepts – Area for Review 1.2 for a summary) would increase, public confidence in the new QCE.

The new QCE system did incorporate these recommendations (see Area for Review 4.1 for a summary). The new QCE system has both Internal and External Assessment activities within each subject. The IA activities now have greater structure and guidance from QCAA (e.g., assessment contents, structure, and activities) yet retain room for teachers to design contextually relevant assessments and use their professional judgements in students' learning. The reliability and validity of the IAs is monitored through an Endorsement, Confirmation and ratification process to increase the comparability of assessments of students across different contexts. There is an External Assessment for each subject that is set and marked by QCAA. Together, these changes to the new QCE system sought to increase comparability and transparency in the assessment system thereby contributing to greater public confidence in the validity and fairness of the QCE.

Public confidence is an important component of any assessment and certification system. This is explicitly recognised in QCAA's own purpose statement, which reads: "To provide high-quality curriculum, assessment and reporting services that enable equity, excellence, public confidence, and lifelong learning and in other assessment programs internationally" (QCAA, 2025g), Public confidence is important for many reasons. It indicates that an assessment or certification system, such as the QCE, is perceived a good indicator of the qualification under consideration (Richardson, 2022) and that it reflects students' achievements in understanding the underlying curriculum.

Public confidence is increased when an assessment system is perceived to be fair, including the extent to which it is reliable, valid and transparent (Ofqual, 2025; QCAA, 2025g; Richardson, 2022). When assessment systems demonstrate consistency (reliability), focus on skills and knowledge relevant to the certification system (validity) and can be generally understood by the public (transparency), they are more likely to be perceived as fair.

In this review, we operationalise public confidence of the new QCE system as the extent to which stakeholders are satisfied with the robustness and fairness of the system. In the following section, we provide a summary of evidence gathered through this evaluation of the public confidence in the new QCE system.

Overall, there are mixed levels of public confidence in the new QCE system. The most positive responses were seen among school leaders and QCAA staff. School leaders agreed that assessment flaws in the old system had been rectified, with 76% agreeing that the new QCE is better than the previous system (see Figure 9.57, Appendix E). Most QCAA staff agreed that the new QCE system is fair and equitable and that it is a high-quality certification system (see Figure 9.58, Appendix E).

This confidence was also reflected in other data sources from the parents' association and local employers. As direct stakeholders in the QCE, their perception of the fairness of the system is a strong reflection of public confidence. Information from the parents' association suggested that parents' confidence was improved when the new system was introduced, particularly because of the inclusion of the external exams. The external exams were noted as a source of public confidence for three reasons.

First, they are a way parents can verify that the work the students and schools are doing is comparable across the system. Second, the exams align with state assessment policies, making parents feel more confident that their children were experiencing a similar educational process to students throughout the state. Third, the invigilation process of the EAs increases perceptions of fairness. In the invigilation survey, nearly all participants suggested that the invigilation process contributes to public confidence in the QCE system (see Figure 9.59, Appendix E). As fairness is a key component of public confidence (Richardson, 2022), the invigilation process would directly contribute to public confidence in the QCE.

This is consistent with data from Queensland employers, many of whom stated that the QCE is one factor they considered when making hiring decisions. Employers interpreted the attainment of a QCE certificate as an indication of effort, persistence and basic academic ability or literacy and numeracy. However, this was not necessarily due to the new system but rather a general understanding of completing Year 12 in either system.

School staff were less confident in the new system. Less than half of respondents agreed or strongly agreed with the statement "the new QCE system is fair and equitable for my students" and less than a quarter agreed or strongly agreed that the new system addresses the needs of community members. Responses to both questions were varied (i.e., somewhat agree

or somewhat disagree) and about a quarter of participants disagreed with these statements (see Figure 9.60, Appendix E). This suggests that work could be done to understand concerns and increase school staff confidence in the system.

Transparency and clear communication are often cited as ways to improve public confidence in assessment systems (Ofqual, 2025; QCAA, 2025g; Richardson, 2022). In addition to providing more information for the schools and the public on the system (e.g., QCAA (2025g)), QCAA may consider following up on these mixed findings by investigating the extent to which stakeholders find the new QCE system transparent and how that relates to their perceptions of fairness of and confidence in the new QCE. QCAA could also explore other elements that might be adjusted to increase confidence.

Findings

- Public confidence has improved since the new system was introduced
- Confidence differs by stakeholder groups: school leaders, parents and employers generally indicated confidence in the system whereas school staff were more moderate

Recommendation

- QCAA consider investigating the extent to which different stakeholders find the new QCE system transparent and how that relates to their perceptions of the fairness of and their confidence in the new QCE

6.5 Academic integrity of the QCE system

Area for review 7.3

Whether the system maintains academic integrity and supports students to act responsibly and ethically.

Changes in the new QCE system were designed to improve the fairness, transparency and comparability of assessments and support students to act responsibly and ethically. Academic integrity refers to students' ethical academic conduct, such as the use of appropriate materials and sources in their work. In this evaluation, we focus on assessment-related changes in the new QCE system and QCAA's resources for academic integrity, as well as potential risk factors to students acting responsibly and ethically during assessments.

As described in Section 6.4 (Area for Review 7.2), the changes to the assessment processes in the new QCE aimed to increase comparability and transparency. These changes included the re-introduction of EAs. These assessments were conducted under standard invigilation procedures, consistent with practices used in previous QCS tests and in other jurisdictions to ensure proper administration and fairness. This process increased stakeholders' confidence in the QCE results, in part due to the standard of academic integrity it helps to uphold (see Area for Review 7.2 for further details).

The new QCE system also included other resources to support academic integrity. The QCE and QCIA Handbook (2025) clearly outlines the responsibilities of schools, teachers, students and parents/carers. It provides examples of academic misconduct and guidance for schools in supporting and promoting academic integrity. In addition, QCAA has developed modules for teachers and students on academic integrity. While the modules are currently voluntary, the student module – taking approximately one hour and consisting of four sections – will become mandatory for all students graduating from Year 12 in 2026 and thereafter, in line with the new QCAA Regulation effective 1 September 2025. Fact sheets on the student module are available for parents/carers (QCAA, 2023b) and teachers (QCAA, 2023c), and students can access the modules through the QCAA portal.

Stakeholder feedback on the modules was generally positive, with some suggestions for improvement to meet student needs. Stakeholders reported that academic integrity had improved in the QCE system, in part due to the modules. However, some noted that the course may be too academic and inaccessible to some students. QCAA could consider reviewing the content and delivery of the modules.

Other evidence suggests that the guidelines and resources are helping students act ethically and responsibly. Data from the invigilator survey show that most students are aware of and uphold academic integrity in the EAs (see Figure 9.61, Appendix E). In addition, invigilators were overwhelmingly satisfied with QCAA's handling of invigilation issues. Together, these findings suggest that QCAA has clear, well-communicated guidelines on academic integrity in assessments that are understood by students.

A major concern in academic integrity is the use of GenAI tools like ChatGPT. Stakeholder data indicate that AI poses a significant risk to academic integrity in the QCE system and education systems globally (Fütterer et al., 2023). AI use is seen as especially problematic for unsupervised IAs. In response, QCAA has updated its guidelines and released resources related to upholding academic integrity. This includes guides for schools (QCAA, 2024g) as well as students and parents/carers (QCAA, 2024h), information brochures (QCAA, 2023a) and webinars (QCAA, 2025a) on AI and academic integrity, and updates to the academic integrity modules. Given growing global concern about AI in learning and assessment, QCAA's proactive response demonstrates its intent to uphold academic integrity in a rapidly changing technology context.

Other academic integrity risks to IAs were also noted. Stakeholders highlighted the use of tutors to complete assessments, a breach that is difficult to identify and prove. They observed that students who could afford tutors may gain an unfair advantage by circumventing academic integrity rules.

Re-using the same IAs each year is not permitted under QCAA policy. However, as in other assessment systems, there remains a risk that newly developed tasks may be similar in scope or structure to those used in previous years, increasing the potential for student plagiarism. This is a challenge for any assessment system that relies on recurring or similar assessments, not just the new QCE (see section 5.8 for further discussion). QCAA seeks to mitigate this risk by requiring assessments to be revised annually, but concerns for plagiarism from previous versions should be considered. AI and students' employment of tutors are considered more significant threats to academic integrity, with no easy solutions. Moving forward, it is important for QCAA to continue exploring strategies to address these risks and uphold academic integrity in the new QCE system.

Findings

- The new QCE system has guidelines and resources to help students understand academic integrity
- AI tools, the employment of tutors and plagiarism are major risks to academic integrity

Recommendation

- QCAA to consider monitoring changes in the distribution of results in Internal and External Assessments over time as an added check on the integrity of Internal Assessments not provided under supervision

6.6 Increased advice to teachers

Area for review 7.4

The result of increasing subject-based and content experts at the QCAA to provide teachers with advice on teaching, learning and assessment.

QCAA has invested significant resources in increasing internal expertise in teaching, learning and assessment, including through the instantiation of a new Research and Innovation Branch. This branch, along with the Senior Curriculum and Assessment Branch, Senior Assessment and Certification Branch and External and Online Assessment Branch (QCAA, 2025f), provides expertise within the agency to support teachers and schools with advice as well as a comprehensive suite of professional development opportunities. Continuing professional development is offered to teachers from K-12, with the focus in Years 11–12 on teaching and assessment resources, particularly assessment literacy (QCAA, undated–a).

QCAA staff members generally favourably rated survey items regarding QCAA's organisational culture and behaviour, especially as it relates to attitudes around monitoring, evaluation and expertise based and evidence informed decision making. However, an item referring to organisational culture around staff learning and development saw 36% of responses in 'disagree' categories. This observation potentially correlates with systemic challenges articulated in the broader report. Sustained organisational pressures arising from continuous systemic transformation and the temporal constraints inherent in the implementation of the new QCE framework have imposed limitations on the support available for staff development and increased internal expertise. Continuing to run any large, complex system on a continuous-change basis is likely to result in burnout of key staff. Regularisation of procedures towards a 'business as usual' approach may make more time and resources available for allocation to internal staff development. Key stakeholders also expressed concerns about the organisational change management, advocating a more structured approach and ensuring sustainable implementation rather than relying on collective goodwill and informal collaborative efforts.

Despite a moderate level of concern regarding staff development, QCAA staff were generally positive and confident in the QCE system and their role in it. QCAA's substantial investment in expertise to develop and keep content up to date in professional development courses, especially as they pertain to assessment literacy, has borne fruit, as can be seen in the survey responses from school staff members. The responses expressed positive views towards their expertise in assessment development and their ability to prepare their students.

The culture that QCAA has built around strong support for educator professional development undergirding high expectations for delivered assessment quality shows clearly in school staff survey items on attitudes towards assessment

quality and approaches. Assessment quality was rated as a high priority, and openness to changes in assessment practice is robust.

Findings

- QCAA has invested significant resources in increasing expertise within the agency
- A moderate level of concern was expressed regarding support for QCAA internal staff development
- QCAA should seek to solidify the new system into a 'business as usual' mode after extensive systemic changes
- QCAA staff felt empowered to enact the new QCE system
- QCAA's resources and professional development are viewed positively by school staff members, QCAA staff and school leaders

6.7 Impact of accreditation and professional development

Area for review 7.5

Whether the new accreditation, training and professional development opportunities have led to an improvement in teachers' expertise, especially in relation to their knowledge of assessment practice and capacity to make high-stakes judgements

The new QCE system and qualification marked a significant change in Queensland's approach to senior assessment and tertiary entrance. To ensure this level of change was appropriately supported, QCAA focused resources on developing teachers' assessment literacy and expertise, particularly for high-stakes judgements. Training and professional development opportunities were central to implementation. These developments are described in detail in section 5.6 and thus will not be extensively detailed here.

QCAA has explicitly acknowledged that "seeking to provide teachers with meaningful professional development that improves their assessment skills and expertise" (QCE & QCIA Handbook, 2025) was a major focus of this effort. The professional development programs were designed to give "schools greater knowledge of the new QCE system", provide "a shared understanding of assessment design and instrument-specific marking guides" and, importantly, to "support schools to improve assessment judgements and report accurate provisional results" (QCAA, 2022b).

The assessment literacy course offered by QCAA consists of seven modules, with a two-stage accreditation process (QCAA, 2021c). Stage 1, launched in January 2018, consists of the first three modules, designed to increase assessment literacy and skills in all senior secondary teachers. Stage 2, released during 2018–2019, consists of four modules, developed for those wishing to become assessors for the QCE system. As of July 2021, more than 110,000 assessment attempts had been recorded for these courses, with the Stage 1 courses having 23,000–33,000 attempts and the Stage 2 courses 5,000–7,000 attempts.

In 2021, QCAA commissioned the Queensland University of Technology (QUT) to complete an evaluation of QCAA's accreditation courses (Willis et al., 2021). QCAA was commended for the high quality of the courses, especially given the intense workload as the new QCE system more broadly was being implemented at the same time. The review made recommendations regarding refreshing content, refreshing the end-of-course assessment, developing additional courses including refresher courses and online recordings of workshops, monitoring and evaluating consequences and increased leadership in messaging regarding benefits of assessment literacy. An updated accreditation program was released by QCAA in July 2022 (QCAA, 2022c). As recommended in the QUT report, QCAA has established agreements with universities to offer credit towards a graduate certificate or master's degree program, increasing the incentive for course completion for teachers pursuing higher education (QCAA, undated–b).

Overall, stakeholders and participants hold positive views of the assessment literacy training. The cumulative evidence offers grounds for optimism that teacher assessment development practices are improving if the higher rates from Endorsement and Confirmation prove to be stable or continues to increase over time. More detail and recommendations are provided in Chapter 5.

Additional data collected through surveys supports these findings. The overall training program received generally positive feedback from all participants (e.g., endorsers, confirmers, invigilators, EA markers, QCAA staff members, school staff members and principals). Survey and interview data suggest the professional development program has been well received. Most teachers, in their roles as markers, confirmers and endorsers, reported that the training has developed their knowledge and capacity to make judgements. Reports from school leaders highlighted the need to continue training new staff, as they perceived newly qualified teachers as lacking the assessment literacy required for the new system.

Findings

- QCAA has developed a suite of assessment literacy training modules that have been widely utilised by Queensland educators
- QCAA has authorised external reviews of the modules and implemented updates based on the findings to keep content current
- QCAA has developed and delivered a set of assessment literacy course modules that are available to Queensland teachers
- Positive views of teacher assessment literacy were found across a range of stakeholders and program participants

6.8 Fairness and accuracy of subject results

Area for review 7.8

The fairness and accuracy of subject results.

6.8.1 Overview

The QCE system's commitment to producing fair and accurate subject results is fundamental to maintaining public confidence and ensuring equitable outcomes for all Queensland students. This analysis examines the fairness and accuracy of subject results through comprehensive psychometric evaluation of nine representative general syllabus subjects (Biology, Business, Design, English, Japanese, Mathematical Methods, Modern History, Physical Education, and Visual Art), integrated with the quality assurance processes detailed in Chapter 5.

The Standards (AERA et al., 2014) provide the internationally accepted benchmark for assessment quality, offering widely recognised criteria for assessment fairness and accuracy that transcend specific system designs. Fairness encompasses multiple dimensions: procedural fairness through consistent application of assessment standards across schools and contexts; distributive fairness ensuring all students have appropriate opportunities to demonstrate their achievement; and outcome fairness whereby results accurately reflect student attainment regardless of background characteristics (AERA et al., 2014). Accuracy refers to the degree to which subject results precisely represent students' true achievement levels, free from measurement error, bias or systematic distortion (AERA et al., 2014).

6.8.2 Understanding QCE subject results and analysis framework

The term "subject results" in the QCE context encompasses multiple interconnected components that require distinct analytical approaches. At the highest level, final subject grades (A–E) represent the ultimate subject result reported to students, derived through algorithmic combination of Internal Assessment scores (typically 75% weighting) and External Assessment scores (typically 25% weighting). These standards-based letter grades represent achievement levels against syllabus objectives and constitute what stakeholders primarily understand as "subject results".

Supporting these final grades are component scores – the underlying numerical scores from individual IAs and EAs that feed into the final grade calculation. These raw scores undergo statistical adjustment before grade determination. At the most granular level, criterion-level performance encompasses the detailed scoring at individual assessment criterion level within IAs, and item-level performance on EAs, which form the foundation of all higher-level scoring.

The psychometric analysis addresses fairness and accuracy across this hierarchical structure through four analytical levels. Criterion/item-level analysis employs Rasch calibration and differential item functioning analysis to examine whether individual assessment components function fairly and accurately across student groups and over time. Component score analysis uses reliability analysis, correlation studies between Internal and External Assessments, and distributional analysis to examine whether the underlying numerical scores accurately reflect student achievement. Grade boundary analysis examines how component scores translate to final letter grades, including whether boundary-setting decisions maintain consistent achievement standards over time. Finally, outcomes analysis investigates potential marking drift and ceiling effects that could compromise the meaning and accuracy of final reported grades.

This multi-level approach recognises that fairness and accuracy of final subject results depend on the integrity of all underlying components, from individual item functioning through to grade boundary placement, all supported by the quality assurance framework described in Section 5.4.

6.8.3 Psychometric evidence of assessment quality

Comprehensive psychometric analysis across 2021–2023, presented in section 9.2 of Appendix B, provides robust evidence regarding the technical quality of QCE assessments. Score reliability across subjects consistently exceeded acceptable standards for high-stakes assessment, with weighted likelihood estimate (WLE) reliability ranging from 0.83 to 0.96. Most IAs achieved reliability above 0.90, with EAs showing similarly strong performance (0.75–0.93). This level of reliability provides confidence that results reflect genuine differences in student achievement rather than measurement error.

Rasch analysis confirmed that assessment criteria within subjects generally function coherently to measure unified constructs. Most items demonstrated acceptable fit to the measurement model, with only scattered instances of poor item fit that, while requiring attention, did not compromise overall assessment integrity. The correlation between Internal and External Assessment ability estimates, ranging from 0.52 to 0.80 across subjects, provides evidence of convergent validity supporting the inference that both components measure related achievement constructs.

However, important variations emerged in how well assessments discriminate across the full range of student ability. While most assessments effectively separated higher and lower-achieving students, several subjects showed suboptimal targeting. In English, Visual Art, and Design, item targeting was limited at both ends of the scale, reducing measurement precision particularly for the lowest and highest-performing students. This creates ceiling and floor effects whereby high-achieving students cannot fully demonstrate their capabilities and low-achieving students cannot effectively enter the assessment. This may compromise fine-grained differentiation of achievement levels.

6.8.4 Assessment fairness across student groups

DIF analysis examined whether assessments function equivalently across different student populations, a critical component of fairness. The comprehensive examination across gender, Aboriginal and Torres Strait Islander status, home language background, school sector, geographical location and AARA status revealed encouraging results.

Very few assessment criteria or items exhibited substantial differential functioning across demographic characteristics. Where moderate DIF was identified, instances were isolated and unlikely to affect overall assessment outcomes significantly. Importantly, no substantial DIF was detected between AARA and non-AARA student groups across any subject, confirming that access arrangements successfully provide equitable assessment conditions without compromising measurement validity.

The temporal stability analysis showed no substantial DIF across the three examination years for Internal Assessment criteria at the individual item level, indicating that specific assessment criteria function consistently in their measurement properties over time. However, this item-level stability must be considered alongside evidence of broader systematic changes in marking patterns, including potential uniform scale drift that is not necessarily detected by DIF analysis. This may be indicative of drift in marking standards in some subjects, with overall performance trends diverging between Internal and External Assessments over the study period.

6.8.5 Quality assurance process effectiveness

The Confirmation process demonstrates effectiveness in maintaining marking consistency, although some areas require attention. Analysis of Confirmation outcomes reveals that minimal percentages of student scores required substantial adjustment (typically less than 1% experiencing changes greater than half a standard deviation), with only Design, Japanese, and Visual Art subjects showing slightly higher rates (2–4%).

The analysis also revealed structural concerns within some subjects' ISMGs that may compromise the reliability of individual criterion judgements. Notably, Biology and Japanese IA1 each contain a single criterion with large rating scales (10 levels in Biology, 15 levels in Japanese), departing from assessment best practice in multiple ways. Having a single criterion reduces assessment reliability compared to multiple independent judgements, while large rating scales show diminishing returns in reliability improvement and impose high cognitive load on markers, potentially leading to inaccurate classifications. These structural limitations in the ISMGs may contribute to the slightly poor model fit observed for IA1C1 in both subjects, and represent systematic constraints on the accuracy of results that extend beyond marking consistency issues.

Additionally, the downward direction of score adjustments suggest the Confirmation process successfully identifies and corrects instances of discrepant marking while maintaining appropriate standards. The small proportion of scores requiring significant adjustment indicates that the Endorsement process effectively prepares schools to apply marking standards consistently, while the Confirmation process provides essential quality control. However, distributional analysis identified potential marking drift in specific subjects, particularly Biology and Mathematical Methods, where IA performance increased more substantially than EA performance over the years examined. This divergence raises concerns about the stability of internal marking standards and suggests the need for enhanced monitoring of marking consistency over time.

6.8.6 Inter-rater reliability and marking consistency

External Assessment marking involving multiple raters (Visual Art, English, Japanese, and Modern History) showed variable inter-rater agreement levels, raising questions about marking consistency. Fleiss' kappa statistics revealed that even when discrepant scores requiring third-marker intervention were excluded, agreement levels ranged from 0.60 to 0.90 across subjects and criteria. While the upper end of this range represents strong reliability, values at the lower end fall below optimal standards for high-stakes assessment, where consistency in marking is critical for fairness.

Moreover, the analysis reveals significant variation in initial marker agreement, with some subjects showing substantial levels of disagreement before third-marker intervention. This pattern indicates that the current marker training and calibration processes, while providing a safety net through the discrepancy resolution system, may be insufficient to ensure the level of consistency expected in high-stakes assessment contexts. The reliance on third-marker intervention in such cases indicates areas where further refinement of training or marking guidelines may be beneficial.

6.8.7 Grade boundary setting and standards maintenance

The QCE system's grade boundary-setting process employing Rasch measurement analysis combined with expert professional judgement generally produces decisions consistent with psychometric evidence. Evaluation across multiple subjects and years demonstrates that expert panels typically set boundaries appropriately reflecting assessment difficulty and student achievement levels.

However, the analysis identified instances where grade boundary decisions were not fully aligned with statistical evidence of assessment difficulty, particularly in subjects where potential marking drift may have influenced expert judgements. The integration of formal equating procedures with expert judgement processes would provide more robust statistical foundations for boundary decisions while preserving the validity of incorporating professional expertise. Equating uses statistical methods to make scores comparable across cohorts and years, while expert judgement contributes contextual understanding of curriculum intent and student performance standards. Combining these approaches would create a more defensible and transparent process for boundary decisions, balancing empirical data with professional expertise.

6.8.8 System-wide monitoring and continuous improvement

The psychometric evidence directly supports the monitoring and review processes outlined in the broader evaluation framework. QCAA's establishment of systematic data collection and analysis procedures enables ongoing quality assurance, although opportunities exist to enhance the proactive use of psychometric data for system improvement.

The current approach, while effectively identifying issues as they arise, could benefit from more systematic predictive monitoring – for example, by analysing trends in endorsement data to identify subjects or schools at risk of recurrent issues, monitoring marker agreement rates over time to detect emerging reliability concerns, or using statistical modelling to flag unexpected shifts in grade distributions before certification. Regular psychometric analysis across all subjects, integrated with established governance structures, would strengthen QCAA's capacity to identify emerging trends and implement preventive measures before they affect student outcomes.

6.8.9 Integration with quality assurance framework

These psychometric findings validate the effectiveness of the quality assurance framework described in Chapter 5 while identifying opportunities for enhancement. The Endorsement process successfully ensures assessment instruments meet syllabus requirements, as evidenced by the strong psychometric properties observed. The Confirmation process effectively maintains marking consistency, though the identification of potential scale drift suggests the need for enhanced monitoring procedures.

The External Assessment component provides reliable common measurement points, although attention to item targeting and inter-rater reliability would strengthen this foundation. The integrated approach combining school-based assessment with external verification creates a robust system for producing fair and accurate results while maintaining the educational benefits of teacher professional judgement. Detailed methodological approaches and, statistical analyses are provided in Appendix B.

Findings

- Most QCE assessments demonstrate strong psychometric properties, with score reliability consistently above acceptable standards (0.83–0.96) and good model fit across most items and criteria
- Assessment criteria function consistently across years with minimal differential item functioning, though some subjects show other evidence of marking drift over time
- DIF analysis revealed minimal bias across demographic characteristics (gender, ATSI status, LBOTE, school sector, geolocation, and AARA), providing evidence of the fairness of the assessments for diverse student populations
- Several subjects show suboptimal targeting of the IAs at the higher achievement levels, creating ceiling and floor effects that limit discrimination among students at the extremes of the performance range and reduce measurement precision
- Some ISMGs have structural limitations impacting reliability, particularly single criteria carrying excessive weight (e.g., 15-mark criteria) and rating scales exceeding optimal cognitive load for markers
- External Assessment marking shows acceptable agreement after resolution processes, though some subjects demonstrate initial disagreement levels that warrant further investigation with respect to marker training
- Moderate to high correlations between IA and EA ability estimates confirm both components measure related achievement constructs

- Endorsement and Confirmation processes do not significantly affect IA criteria performance. A very small percentage of students experience significant score changes during Confirmation
- Grade boundary-setting processes are generally sound, with expert judgements mostly consistent with statistical evidence, though misalignment exists for some subjects at key boundaries

Recommendations

- The QCAA Board to establish a Research Advisory Committee with a majority of external members, including a smaller, specialised group to deal with technical issues
- QCAA to investigate alignment between assessment difficulty and student ability to maximise precision and eliminate ceiling and floor effects
- QCAA to establish a common measurement scale across years to enable monitoring of assessment targeting and student progress over time
- QCAA to enhance the psychometric monitoring procedures integrated within the existing quality assurance processes, including enhanced linking approaches and differential item/test functioning analyses, particularly to monitor and address potential marking drift and the functioning of assessment criteria over time
- QCAA to conduct a systematic review of ISMGs using psychometric evidence to revise under-performing criteria, particularly single-criterion assessments and excessively large rating scales
- QCAA to implement enhanced marker training and calibration for EA subjects showing persistent elevated initial disagreement levels, with ongoing monitoring and rapid intervention protocols
- QCAA to combine statistical evidence (equating, predictive models, sequential validation) with expert judgement when setting grade boundaries, following international best practice
- QCAA to develop comprehensive documentation of boundary-setting processes and implement periodic external reviews by independent measurement experts

07
Conclusion

7. Conclusion

This evaluation considered the effectiveness of the development and implementation of the new QCE system and the delivery of equitable outcomes for students. It evaluated the current processes and made recommendations for future improvements.

The evaluation methods encompassed a comprehensive approach, drawing on multiple sources of evidence. These included stakeholder consultations, in-depth analysis of a sample of student outcomes and performance data provided by QCAA, direct observations of key assessment processes and activities, targeted surveys and interviews with diverse stakeholders such as students, teachers, parents, school leaders, QCAA staff and associates, and comprehensive psychometric analyses of the assessment data for a representative subset of subjects. To enhance the rigour of the findings, the evaluation incorporated a targeted review of relevant literature and benchmarking against established best practice criteria, thereby ensuring a robust and triangulated evaluation.

It found that QCAA has overseen a significant reform of the QCE system, with strong positive feedback from stakeholders on transparent communication during development and thorough trialling processes. These efforts supported a smooth transition to the revised QCE procedures. In several areas, the system aligns with international best practice, including systemwide capacity-building and professional practice. QCAA has achieved this through the systemic implementation of Endorsement, Confirmation, and External Assessment.

The implementation of the new complex system throughout the state of Queensland has been successful. Fidelity with the intended system design also appears to be effective across the state. The new portal and custom software applications have ensured secure access to information and improvement. Information on AARA is comprehensive and easily accessed online, and no indications of bias for students with AARA was found.

This comprehensive review of the new QCE system reveals a sophisticated and forward-looking approach to educational assessment in Queensland which continues to emphasise trust in teachers as assessors. QCAA has developed a unique system that strategically balances regulatory oversight with institutional autonomy, underpinned by a comprehensive quality assurance framework. The design intentionally sought to balance local contextualisation of learning and assessment with the maximisation of reliability and state-wide comparability of student results. The latter is facilitated by the use of EAs that ensure a high degree of academic rigor and the comparability of students across contexts. The former is integrated through syllabuses that clearly delineate the required content while leaving the specific pedagogical approach to teachers. Syllabuses focus on cognitive complexity and higher-order thinking, using a cognitive verb framework to promote critical and analytical thinking. The syllabuses are standards-based and specify detailed achievement benchmarks, objectives, and assessment criteria aligned with senior secondary schooling outcomes. Teacher professional input and feedback contribute to syllabus development, review and revisions.

An important strength of the current system is the stakeholder engagement, as documented in extensive consultations and transparent communication with a diverse group of stakeholders throughout the implementation processes. Further, quality assurance includes rigorous processes such as Endorsement, Confirmation and External Assessments, where QCAA has established targeted professional development throughout the implementation phase grounded in local contextual understanding. Psychometric analyses consistently demonstrate reliable and valid assessment results with minimal demographic bias. Moreover, the incorporation of learning beyond traditional classroom contexts as an option was found to be valuable. However, there is a need to review the credit awarded for VET Certificate and Diploma courses through comparison with subjects with syllabuses developed by QCAA.

While stakeholders broadly express satisfaction with the system's flexibility and core components, several areas require attention to strengthen confidence, drive continuous improvement, and ensure long-term sustainability. Key areas for improvement include enhancing communication channels to ensure universal access to information, addressing concerns about decision consistency, and refining syllabus rollout timing. Emerging challenges are also evident, particularly regarding academic integrity in the context of advancing technologies.

Moving forward, the QCE system is well placed to tackle some of the challenges emerging from the changing educational landscape globally. As with any complex system comprising multiple processes and components and numerous participants and stakeholders, the evaluation identified opportunities to further improve the QCE system design and delivery. These include: developing a comprehensive research agenda to identify factors that might inform revision or adjustment of current syllabuses, with a focus on monitoring implementation and its impact on delivery conditions, timelines, and teacher and assessor workload; expanding external representation in QCAA governance; establishing a measurement advisory board with an emphasis on providing guidance on technical and psychometric issues; investigating the extent to which experiences with the AARA process vary across different student or school factors; reviewing the assessment literacy modules offered to teachers to ensure they are providing the most effective and up-to-date guidance for educators; increasing use of psychometric analyses to monitor the new QCE system and maintain public confidence in the fairness and accuracy of subject results; augmenting the monitoring and quality assurance of marking; enhancing the processes used in grade-boundary setting with statistical evidence and clearer documentation; and improving targeting of assessments to assure sufficient coverage of the full ability range of student performances. Some critical areas the evaluation has identified are related to technological integration. It is recommended to explore advanced technologies like generative AI and large language models to streamline procedures while maintaining human expertise in critical decision-making.

The improvements would be facilitated by reconfiguring the QCAA Board to include more external members, broadening the perspectives, advice, and support they can offer the agency; instantiating a Research Advisory Committee with a majority of external and independent members; and creating a specialised advisory group focused on technical, statistical, and psychometric issues. Such committees could further be part of the development of a comprehensive research agenda, to monitor implementation, assess impact and proactively address emerging challenges, particularly in the areas of academic integrity and technological advancement and to monitor possible variations in assessment experiences across different student and school contexts to ensure universal accessibility and fairness for all.

To conclude, the QCE system is well-positioned to support the Queensland's educational landscape, aligning with international best practices while maintaining a commitment to innovation and continuous improvement. Ongoing modernisation of assessment processes, combined with a strategic approach to incorporating new techniques and tools, will be crucial in supporting Queensland's continued development within Australian and global secondary education assessment frameworks in the future. By embracing emerging technologies, maintaining rigorous quality standards, and remaining responsive to stakeholder feedback, QCAA can ensure the long-term sustainability and effectiveness of the QCE system.

08
References

8. References

ACARA. (2020). *The Shape of the Australian Curriculum Version 5*. https://www.acara.edu.au/docs/default-source/curriculum/the_shape_of_the_australian_curriculum_version5_for-website.pdf

ACARA. (2025). *Assessment*. www.acara.edu.au/assessment

ACARA. (undated–a). *The Australian Curriculum: Version 9.0*. https://www.australiancurriculum.edu.au/

ACARA. (undated–b). *NAPLAN*. https://www.nap.edu.au/naplan

AERA, APA, & NCME. (2014). *Standards for educational and psychological testing* (2014 English edition. ed.). American Educational Research Association.

Andersson, B., Bränberg, K., & Wiberg, M. (2013). Performing the Kernel Method of Test Equating with the Package kequate. *Journal of Statistical Software*, *55*(6). https://doi.org/10.18637/jss.v055.i06

Aston, R., Paproth, H., Robertson, P., Sun, S. Z., Luo, R., McGrane, J., Zhang, Z., & Hopfenbeck, T. N. (2024). *Evaluation Framework for the new QCE System: prepared for QCAA*. University of Melbourne.

Australian Government. (1992). *Disability Discrimination Act 1992*. https://www.legislation.gov.au/C2004A04426/2018-04-12/text

Australian Government. (2005). *Disability Standards for Education 2005*. https://www.education.gov.au/disability-standards-education-2005

Baird, J.-A., Hopfenbeck, T. N., Elwood, J., Caro, D., & Ahmed, A. (2014). *Predictability in the Irish Leaving Certificate*.

Bardhan, P. (2002). Decentralization of governance and development. *Journal of Economic Perspectives*, *16*(4), 185–205. https://doi.org/10.1257/089533002320951037

Bennett, R. E. (2024). Personalizing assessment: Dream or nightmare? *Educational Measurement, Issues and Practice*, *43*(4), 119–125. https://doi.org/10.1111/emip.12652

Berman, A. I., Haertel, E. H., & Pellegrino, J. W. (2020). *Comparability of large-scale educational assessments: Issues and recommendations*. National Academy of Education.

Blacher, Y. (2025). *Organisational review of the Victorian Curriculum and Assessment Authority (VCAA)*. https://www.education.vic.gov.au/Documents/Stage-2-Organisational-Review-of-the-VCAA.pdf

Black, P. (2016). The role of assessment in pedagogy–and why validity matters. In D. Wyse, L. Hayward, & J. Pandya (Eds.), *Sage handbook of curriculum, pedagogy and assessment* (Vol. 2, pp. 725–739). Sage.

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, *25*(6), 551–575. https://doi.org/10.1080/0969594x.2018.1441807

Board of Senior Secondary School Studies. (1978). *A Review of school-based assessment (ROSBA) in Queensland secondary schools*.

Board of Senior Secondary School Studies. (2011). *Review of school-based assessment (ROSBA): discussion papers*. https://www.qcaa.qld.edu.au/downloads/publications/research_qbssss_rosba_11.pdf

Burns, T., Köster, F., & Organisation for Economic Co-operation and Development. (2016). *Governing education in a complex world*. OECD Publishing.

Capano, G. (2011). Government continues to do its job. A comparative study of governance shifts in the higher **education sector**. *Public Administration*, *89*(4), 1622–1642. https://doi.org/10.1111/j.1467-9299.2011.01936.x

CAST. (2024). *The UDL Guidelines*. https://udlguidelines.cast.org/

Choi, B. C., & Pak, A. W. (2005). A catalog of biases in questionnaires. Preventing chronic disease. *Preventing chronic disease*, *2*(1), A13.

Clarke, E. (1987). *Assessment in Queensland secondary schools: Two decades of change 1964-1983. Historical perspectives on contemporary Issues in Queensland education No. 4*. Queensland Department of Education.

Clarke, E. (1990). *Assessment in Queensland secondary schools: 1983-1990*. Queensland Department of Education.

Constantinou, F. (2024). 'If you have a question that doesn't work, then it's clearly going to upset candidates': What gives rise to errors in examination papers? *Oxford Review of Education*, *51*(3), 357–377. https://doi.org/https://doi.org/10.1080/03054985.2024.2308548

Cumming, J., Dickson, E., & Webster, A. (2013). Reasonable adjustments in assessment: Putting law and policy into practice in Australia. *International Journal of Disability, Development and Education*, *60*(4), 295–311. https://doi.org/10.1080/1034912X.2013.846467

Darling-Hammond, L. (2009). *The flat world and education: How America's commitment to equity will determine our future*. Teachers College Press.

Darling-Hammond, L., Newton, S. P., & Wei, R. C. (2013). Developing and assessing beginning teacher effectiveness: The potential of performance assessments. *Educational Assessment, Evaluation and Accountability*, *25*(3), 179–204. https://doi.org/10.1007/s11092-013-9163-0

DeLuca, C. (2012). Preparing teachers for the age of accountability: Toward a framework for assessment education. *Action in Teacher Education*, *34*(5-6), 576–591. https://doi.org/10.1080/01626620.2012.730347

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Approaches to classroom assessment inventory: A new instrument to support teacher assessment literacy. *Educational Assessment*, *21*(4), 248–266. https://doi.org/10.1080/10627197.2016.1236677

Elliott, S. N., & Marquart, A. M. (2004). Extended time as a testing accommodation: Its effects and perceived consequences. *Exceptional Children*, *70*(3), 349–367. https://doi.org/10.1177/001440290407000306

ETS. (2023). *Pool of analytical writing topics*. https://www.ets.org/pdfs/gre/analytical-writing-pool.pdf

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(23), 8410–8415.

Fütterer, T., Fischer, C., Alekseeva, A., Chen, X., Tate, T., Warschauer, M., & Gerjets, P. (2023). ChatGPT in education: Global reactions to AI innovations. *Scientific Reports*, *13*(1), 1–14. https://doi.org/10.1038/s41598-023-42227-6

Ganzfried, S., & Yusuf, F. (2018). Optimal weighting for exam composition. *Education Sciences*, *8*(1), 36. https://doi.org/https://doi.org/10.3390/educsci8010036

Gore, J., Ellis, H., Fray, L., Smith, M., Lloyd, A., Berrigan, C., Lyell, A., Weaver, N., & Holmes, K. (2017). *Choosing VET: Investigating the VET aspirations of school students*.

Grek, S. (2015). Seeing from the top of the tower: PISA and the new governing panoramas in Europe. *Compare*, *45*(3), 479–481. https://doi.org/10.1080/03057925.2015.1027512

Gwet, K. L. (2019). *irrCAC: Computing Chance-Corrected Agreement Coefficients (CAC)*. In (Version 1.0) https://cran.r-project.org/web/packages/irrCAC/irrCAC.pdf

Hamilton, M., Maddox, B., & Addey, C. (2015). *Literacy as numbers: Researching the politics and practices of international literary assessment*. Cambridge University Press.

He, Y., Cui, Z., & Osterlind, S. J. (2015). New Robust Scale Transformation Methods in the Presence of Outlying Common Items. *Applied Psychological Measurement*, *39*(8), 613–626. https://doi.org/10.1177/0146621615587003

Hopfenbeck, T. N. (2016). The power of PISA – limitations and possibilities for educational research. https://doi.org/https://doi.org/10.1080/0969594X.2016.1247518

Hopfenbeck, T. N., Flórez Petour, M. T., & Tolo, A. (2015). Balancing tensions in educational policy reforms: Large-scale implementation of assessment for learning in Norway. *Assessment in Education: Principles, Policy & Practice*, *22*(1), 44–60. https://doi.org/10.1080/0969594X.2014.996524

Hopfenbeck, T. N., & Görgen, K. (2017). The politics of PISA: The eedia, policy and public responses in Norway and England. *European Journal of Education*, *52*(2), 192–205. https://doi.org/10.1111/ejed.12219

Hopfenbeck, T. N., Tolo, A., Florez, T., & El Masri, Y. (2013). Balancing trust and accountability? The assessment for learning programme in Norway: A governing complex education systems case study. *OECD Education Working Papers*, *97*. https://doi.org/https://doi.org/10.1787/5k3txnpqlsnn-en

Hunter, N. (2011). The QCE and the three Rs: relevance, rigour and reputation.

Jenkins, G. (2020). Teacher agency: The effects of active and passive responses to curriculum change. *Australian Educational Researcher*, *47*(1), 167–181. https://doi.org/10.1007/s13384-019-00334-2

King, E. M., & Özler, B. (2005). What's decentralization got to do with learning? School autonomy and student performance in Nicaragua's reform. *World Bank Policy Research Working Paper No. 3544*. https://doi.org/https://doi.org/10.1596/1813-9450-3544

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). Springer.

Koretz, D. M. (2008). *Measuring up: What educational testing really tells us*. Harvard University Press.

Laitusis, C. C., & Karvonen, M. (2025). A case for reimagining Universal Design of assessment systems. *Educational Measurement, Issues and Practice*. https://doi.org/https://doi.org/10.1111/emip.12674

Lane, S., Raymond, M. R., & Haladyna, T. M. (2015). Test development process. In M. R. R. S. Lane, & T. M. Haladyna (Ed.), *Handbook of test development* (pp. n.p.). Routledge.

Lee, S. Y., Lee, J. C.-K., & Lam, B. Y.-H. (2022). Does renaming improve public attitudes toward vocational education and training in higher education? Evidence from a survey experiment. *Education + Training*, *64*(3), 347–359. https://doi.org/https://doi.org/10.1108/ET-01-2021-0014

Maclellan, E. (2004). Initial knowledge states about assessment: Novice teachers' conceptualisations. *Teaching and Teacher Education*, *20*(5), 523–535. https://doi.org/10.1016/j.tate.2004.04.008

Marzano, R. J., & Kendall, J. S. (2007). *The new taxonomy of educational objectives* (2nd ed.). Hawker Brownlow Education.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149–174. https://doi.org/10.1007/bf02296272

Matters, G., & Masters, G. N. (2014). *Redesigning the secondary-tertiary interface: Queensland Review of Senior Assessment and Tertiary Entrance*. Australian Council for Educational Research (ACER).

McCaffrey, D. F., Casabianca, J. M., Ricker-Pedley, K. L., Lawless, R. R., & Wendler, C. (2022). Best Practices for Constructed-Response Scoring. Research Report. ETS RR-22-17. *ETS Research Report Series*.

McGaw, B. (1977). Ought secondary exit to serve tertiary entrance? Measurement alternatives to the Higher School Certificate. *Bulletin of the Victorian Institute for Educational Research*, *39*, 2–24.

McGaw, B. (1997). *Shaping their future: Recommendations for reform of the Higher School Certificate*. Dept. of Training and Education Co-ordination N.S.W.

McGaw, B., & Hannan, B. (1985). *Certification in upper secondary education*. Curriculum Development Centre.

McGaw, B., Warry, R. S., & McBryde, B. (1975). Validation of an aptitude measure for the rescaling of school assessments. *Education Research and Perspectives*, *2*, 20–34.

McLean, P., Perkins, K., Tout, D., Brewer, K., & Wyse, L. (2012). Australian Core Skills Framework (ACSF). *Department of Industry, Innovation, Science, Research and Tertiary Education.* https://doi.org/https://research.acer.edu.au/transitions_misc/12

Mikkonen, S., Pylvas, L., Rintala, H., Nokelainen, P., & Postareff, L. (2017). Guiding workplace learning in vocational education and training: a literature review. *Empirical Research in Vocational Education and Training*, *9*. https://doi.org/10.1186/s40461-017-0053-4

National Center for Education Statistics. (undated). *NAEP assessment process*.

NCCD. (2025). *Nationally consistent collection of data on school students with disability guidelines: Effective 2026 onward*. https://www.nccd.edu.au/sites/default/files/NCCD%20Guidelines%20Effective%202026%20Onward.pdf

Newton, P. (1997). Examining Standards Over Time. *Research Papers in Education: Policy and Practice*, *12*(3), 227–247.

OECD. (1999). *Measuring student knowledge and skills: A new framework for assessment*. OECD.

OECD. (2001). *Knowledge and skills for life: First results from the OECD Programme for International Student Assessment (PISA) 2000*. Organisation for Economic Co-operation and Development.

OECD. (2016). *Education at a Glance 2016: OECD Indicators*. OECD Publishing. https://doi.org/10.187/eag-2016-en

OECD. (2018). *Education at a Glance 2018*. OECD Publishing.

OECD. (2019). *OECD Future of education and skills 2030: OECD Learning Compass 2030*. https://www.oecd.org/content/dam/oecd/en/about/projects/edu/education-2040/1-1-learning-compass/OECD_Learning_Compass_2030_Concept_Note_Series.pdf

OECD. (2020). *What Students Learn Matters: Towards a 21st century curriculum*.

OECD. (2023). *PISA 2025 Science Framework (Draft)*. OECD Publishing.

OECD. (2024). *PISA 2022 Technical Report*. OECD Publishing. https://doi.org/10.1787/01820d6d-en

OECD. (2025a). *Learning in the digital world*. https://www.oecd.org/en/topics/sub-issues/learning-in-the-digital-world/pisa-2025-learning-in-the-digital-world.html

OECD. (2025b). *PISA foreign language assessment*. https://www.oecd.org/en/topics/sub-issues/foreign-language-learning/pisa-2025-foreign-language-assessment.html

OECD/CERI. (1990). *Curriculum Reform: An Overview of Trends.* (O. Publishing, Ed.).

OECD/CERI. (1994). *The curriculum redefined: Schooling for the 21st century* (O. Publishing, Ed.).

Ofqual. (2025). *Our governance*. www.gov.uk/government/organisations/ofqual/about/our-governance

Paek, I., & Wilson, M. (2011). Formulating the Rasch differential item functioning model under the marginal maximum likelihood estimation context and its comparison with Mantel-Haenszel procedure in short test and small sample conditions. *Educational and Psychological Measurement*, *71*(6), 1023–1046. https://doi.org/10.1177/0013164411400734

Page, A., Barr, M., Rendoth, T., Roche, L., Foggett, J. L., Leonard, C., & Duncan, J. (2024). Making reasonable adjustments for students with disability in Australian mainstream classrooms: A scoping review. *Australasian Journal of Special and Inclusive Education*, *48*(1), 46–63. https://doi.org/10.1017/jsi.2024.1

Pitman, J., Herschell, P., Allen, R., Veerman, M., Gray, K., Harris, K., Bell, E., Kelly, D., Brennan, P., & Delley, D. (2002). *The Senior Certificate: A new deal*.

Phi Delta Kappan. (2022). *2022 PDK Poll of the Public's Attitudes Toward the Public Schools: Topline survey results* (54th annual). Phi Delta Kappan / PDK Poll.

Powers, D. E. (2005). Effects of preexamination disclosure of essay prompts for the GRE analytical writing assessment. *ETS Research Report Series*, *2005*(1), 1–21. https://doi.org/https://doi.org/10.1002/j.2333-8504.2005.tb01978.x

QCAA. (2017). *The new QCE system: Design and development strategy*. https://www.qcaa.qld.edu.au/downloads/senior/new_snr_assess_qce_design_develop_strategy.pdf

QCAA. (2018). *QCE literacy and numeracy requirements*. https://www.qcaa.qld.edu.au/downloads/senior/snr_new_assess_te_qce_factsheet_lit_num.pdf

QCAA. (2020). *Guideline for calculating General subject results*. https://www.qcaa.qld.edu.au/downloads/senior/snr_calc_subject_results_legislative_guide.pdf

QCAA. (2021a). *Annual report 2020-2021*. https://www.qcaa.qld.edu.au/downloads/publications/qcaa_annual_report_2021.pdf

QCAA. (2021b). *Attributes of quality assessment*. https://www.qcaa.qld.edu.au/about/k-12-policies/student-assessment/understanding-assessment/attributes-quality-assessment

QCAA. (2021c). *New Queensland Certificate of Education system post-cycle process review report*. https://www.qcaa.qld.edu.au/downloads/about/qce_post_cycle_process_review.pdf

QCAA. (2022a). *Embedding the general capabilities*. https://www.qcaa.qld.edu.au/downloads/aciq/general-resources/ac_gc_embedding_general_capabilities.pdf

QCAA. (2022b). *The QCE Assessment literacy program*. https://www.qcaa.qld.edu.au/senior/assessment/qcaa-assessors/qce-assessment-literacy-program

QCAA. (2022c). *Release of updated accreditation program. Memo number 035/22*. https://www.qcaa.qld.edu.au/downloads/memos/22/035-22.pdf

QCAA. (2023a). *Academic integrity and artificial intelligence (AI)*. https://www.qcaa.qld.edu.au/copyright

QCAA. (2023b). *Academic integrity course for students completing senior studies: Information for parents/carers*. https://www.qcaa.qld.edu.au/downloads/senior/snr_academic_integrity_factsheet_parents_carers.pdf

QCAA. (2023c). *Academic integrity course for students completing senior studies: Information for teachers*. https://www.qcaa.qld.edu.au/downloads/senior/snr_academic_integrity_factsheet.pdf

QCAA. (2023d). *Completing an AARA application: Access arrangements and reasonable adjustments*. https://www.qcaa.qld.edu.au/downloads/senior/aara_completing_application.pdf

QCAA. (2023e). *Credit and learning*. https://www.qcaa.qld.edu.au/senior/certificates-and-qualifications/qce/credit-learning

QCAA. (2023f). *Illness and misadventure: Access arrangements and reasonable adjustments*. https://www.qcaa.qld.edu.au/downloads/senior/aara_illness_misadventure.pdf

QCAA. (2023g). *Imputed disability: Access arrangements and reasonable adjustments*. https://www.qcaa.qld.edu.au/downloads/senior/aara_imputed-disability_factsheet.pdf

QCAA. (2023h). *Senior Education and Training (SET) Plans*. https://education.qld.gov.au/careers/set-plan

QCAA. (2023i). *Sorry business and sad news: Access arrangements and reasonable adjustments*. https://www.qcaa.qld.edu.au/downloads/senior/aara_sorry_bus_sad_news.pdf

QCAA. (2023j). *Student learning accounts*. https://www.qcaa.qld.edu.au/senior/certificates-and-qualifications/slaq

QCAA. (2024a). *21st century skills for senior curriculum. A position paper.* https://www.qcaa.qld.edu.au/downloads/senior/snr_syll_redev_21st_century_skills_position_paper.pdf

QCAA. (2024b). *Access arrangements & reasonable adjustments (AARA)*. https://www.qcaa.qld.edu.au/senior/assessment/aara

QCAA. (2024c). *Annual report 2023-2024*. https://www.qcaa.qld.edu.au/downloads/publications/qcaa_annual_report_2324.pdf

QCAA. (2024d). *Biology 2025 v1.2 General senior syllabus*. https://www.qcaa.qld.edu.au/downloads/senior-qce/syllabuses/snr_biology_25_syll.pdf

QCAA. (2024e). *Business 2025 v1.2 General senior syllabus*. https://www.qcaa.qld.edu.au/downloads/senior-qce/syllabuses/snr_business_25_syll.pdf

QCAA. (2024f). *Design 2025 v1.2 General senior syllabus*. https://www.qcaa.qld.edu.au/downloads/senior-qce/syllabuses/snr_design_25_syll.pdf

QCAA. (2024g). *Developing artificial intelligence capabilities: Guidance for schools*. https://www.qcaa.qld.edu.au/downloads/p_10/factsheet_ai_guidance_schools.pdf

QCAA. (2024h). *Developing artificial intelligence capabilities: Guidance for students and parents/carers*. https://www.qcaa.qld.edu.au/downloads/p_10/factsheet_ai_guidance_students.pdf

QCAA. (2024i). *Endorser manual*. https://www.qcaa.qld.edu.au/downloads/portal/endorsement/endorse_endorser_manual.pdf

QCAA. (2024j). *English 2025 v1.2 General senior syllabus*. https://www.qcaa.qld.edu.au/downloads/senior-qce/syllabuses/snr_english_25_syll.pdf

QCAA. (2024k). *Japanese 2025 v1.2 General senior syllabus*. https://www.qcaa.qld.edu.au/downloads/senior-qce/syllabuses/snr_japanese_25_syll.pdf

QCAA. (2024l). *Mathematical Methods 2025 v1.2 General senior syllabus*. https://www.qcaa.qld.edu.au/downloads/senior-qce/syllabuses/snr_maths_methods_25_syll.pdf

QCAA. (2024m). *Modern History 2025 v1.2 General senior syllabus*. https://www.qcaa.qld.edu.au/downloads/senior-qce/syllabuses/snr_modern_history_25_syll.pdf

QCAA. (2024n). *Physical Education 2025 v1.2 General senior syllabus*. https://www.qcaa.qld.edu.au/downloads/senior-qce/syllabuses/snr_pe_25_syll.pdf

QCAA. (2024o). *Visual Art 2025 v1.2 General senior syllabus*. https://www.qcaa.qld.edu.au/downloads/senior-qce/syllabuses/snr_visual_art_25_syll.pdf

QCAA. (2025a). *Academic integrity and artificial intelligence webinar*. https://www.qcaa.qld.edu.au/about/k-12-policies/artificial-intelligence/webinar-series/academic-integrity-artificial-intelligence

QCAA. (2025b). *Eligibility for a QCIA*. https://www.qcaa.qld.edu.au/senior/certificates-and-qualifications/qce-qcia-handbook/3-qcia/3.1-eligibility

QCAA. (2025c). Endorsement outcomes by event, 2019-24 [unpublished data].

QCAA. (2025d). *Governance*. https://www.qcaa.qld.edu.au/about/governance/members

QCAA. (2025e). *myQCE*. https://myqce.qcaa.qld.edu.au/

QCAA. (2025f). *QCAA organisational structure as at 28 February 2025*. https://www.qcaa.qld.edu.au/downloads/about/qcaa_org_chart.pdf

QCAA. (2025g). *Strategic plan 2023-27*. https://www.qcaa.qld.edu.au/downloads/about/qcaa_strategic_plan_2023-27.pdf

QCAA. (2025h). *Understanding K-12 assessment: K-12 policies and resources*. https://www.qcaa.qld.edu.au/downloads/aciqv9/general-resources/assessment/ac9_understanding-_k-12_assessment.pdf

QCAA. (undated–a). *Continuing professional development for teachers*. https://www.qcaa.qld.edu.au/pd-events/cpd-teachers

QCAA. (undated–b). *The QCE Assessment literacy program*. https://www.qcaa.qld.edu.au/senior/assessment/qcaa-assessors/qce-assessment-literacy-program

QCE & QCIA Handbook. (2025). *QCE and QCIA policy and procedures handbook v6.0*. Queensland Curriculum & Assessment Authority. https://www.qcaa.qld.edu.au/senior/certificates-and-qualifications/qce-qcia-handbook

QHRC. (2019). *The Queensland Human Rights Act*. https://www.qhrc.qld.gov.au/__data/assets/pdf_file/0005/25448/QHRC_factsheet_HRA_easyreadguide.pdf

Queensland Government. (2005). *Education (Queensland College of Teachers) Act 2005*. https://www.legislation.qld.gov.au/view/html/inforce/current/act-2005-047

Queensland Government. (2006). *Education (General Provisions) Act 2006*. https://www.legislation.qld.gov.au/view/html/inforce/current/act-2006-039

Queensland Government. (2014). *Education (Queensland Curriculum and Assessment Authority) Act 2014*. https://www.legislation.qld.gov.au/view/html/inforce/current/act-2014-003

Queensland Government. (2017). *Education (Accreditation of Non-State Schools) Act 2017*. https://www.legislation.qld.gov.au/view/html/inforce/current/act-2017-024

Queensland Government. (undated). *Advancing futures: New senior assessment and tertiary entrance systems in Queensland*.

R Core Team. (2024). *R: A language and environment for statistical computing*. In https://www.R-project.org/

Radford Committee. (1970). *Public Examinations for Queensland Secondary School Students: Report (Radford Report)*. Queensland Department of Education.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danmarks Paedagogiske Institut.

Research and Curriculum Branch. (1972). *The use of school assessments scaled against the Australian Scholastic Aptitude Test for compilation of an order of merit list*. Queensland Department of Education.

Reynolds, R. M. (2004). Vocational Education and the Great Divide: Have student needs been overlooked? *Journal of Educational Research & Policy Studies*, *4*(1), 1–14.

Richardson, M. (2022). *Rebuilding public confidence in educational assessment*. UCL PRESS.

Robitzsch, A. (2020). Robust Haebara linking for many groups in the case of partial invariance. *Psych*, *2*, 155–173. https://doi.org/doi:10.3390/psych2030014

Robitzsch, A., Kiefer, T., & Wu, M. (2025). *Test analysis modules*. In (Version 4.2-21) https://github.com/alexanderrobitzsch/TAM

Sadler, D. R. (1989). Formative Assessment and the Design of Instructional-Systems. *Instructional Science*, *18*(2), 119–144. https://doi.org/10.1007/Bf00117714

Saguin, K. I. (2019). Designing effective governance of education. *Policy Design and Practice*, *2*(2), 182–197. https://doi.org/10.1080/25741292.2019.1621034

Sahlberg, P. (2016). The global educational reform movement and its impact on schooling. In *Handbook of Global Education Policy* (pp. 128–144). https://doi.org/10.1002/9781118468005

Schleicher, A. (2009). International benchmarking as a lever for policy reform. In A. Hargreaves & M. Fullan (Eds.), *Change wars* (pp. 97–115). Solution Tree.

School Curriculum and Standards Authority. (2025). *WACE Manual 2025: General information for senior secondary schooling*. https://scsa.wa.edu.au/media/documents/publications/WACE-Manual_2025.PDF

Shergold, P. R., Calma, T., Russo, S., Walton, P., Westacott, J., Zoellner, D., & O'Reilly, P. (2020). *Looking to the future: Report of the review of senior secondary pathways into work, further education and training*.

Sinharay, S. (2019). Added value of subscores and hypothesis testing. *Journal of Educational and Behavioral Statistics*, *44*(1), 25–44. https://doi.org/https://doi.org/10.3102/1076998618788862

Soguel, N. C., & Jaccard, P. (Eds.). (2008). *Governance and performance of education systems*. Springer.

Stabback, P. (2016). *What makes a quality curriculum?* (Current and Critical Issues in Curriculum and Learning, Issue.

Stanley, J. C., & Wang, M. D. (1970). Weighting test items and test-item options: An overview of the analytical and empirical literature. *Educational and Psychological Measurement*, *30*, 21–25.

Sweller, J., Ayres, P. L., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.

Teo, I., Hsien, M., Buckley, S., & Roberts, A. (2023). *International comparison of tertiary education systems*. https://research.acer.edu.au/cgi/viewcontent.cgi?article=1083&context=higher_education

Thompson, S., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments (Synthesis Report 44)*. ational Center on Educational Outcomes

Tindall-Ford, S., Agostinho, S., & Sweller, J. (Eds.). (2020). *Advances in cognitive load theory: Rethinking teaching*. Routledge.

Tyler, R. W. (1949). *Basic principles of curriculum and instruction*. University of Chicago Press.

UNCRC. (1989). *Convention on the Rights of the Child*. https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-child

van den Akker, L., Heres, L., Lasthuizen, K., & Six, F. E. (2009). Ethical leadership and trust: It's all about meeting expectations. *International Journal of Leadership Studies*, *5*(2), 102–122.

VCAA. (2025). *VCE Administrative Handbook 2025*. https://www.vcaa.vic.edu.au/administration/vce-administrative-handbook/vce-administrative-handbook-2025

Victorian Government. (2025). *Study options in year 11 and 12*. https://www.vic.gov.au/study-options-year-11-and-12

Viviani, N. (1990). *The review of tertiary education in Queensland 1990*. Queensland Department of Education.

Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development [Abstract]. *The Canadian journal of education*, *30*, 751–772. https://doi.org/10.2307/20466661

von Davier, A. A., Holland, P. W., & Thayer, D. T. (2004). *The Kernel method of test equating*. Springer.

Voogt, J. M., Pieters, J. M., & Handelzalts, A. (2016). Teacher collaboration in curriculum design teams: Effects, mechanisms, and conditions. *Educational Research and Evaluation*, *22*(3-4), 121–140. https://doi.org/10.1080/13803611.2016.1247725

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wang, M. W. (1998). Differential weighting: A review of methods and empirical studies. *Review of Educational Research*, *40*, 663–705.

Wei, X., & Zhang, S. (2024). Extended time accommodation and the academic, behavioral, and psychological outcomes of students with learning disabilities. *Journal of Learning Disabilities*, *57*(4), 242–254. https://doi.org/10.1177/00222194231195624

Wiberg, M., & Bränberg, K. (2015). Kernel equating under the non-equivalent groups with covariates design. *Applied Psychological Measurement*, *39*(5), 349–361. https://doi.org/10.1177/0146621614567939

West, M. R. (2014, October 23). *Why do Americans rate their local public schools so favorably?* Brown Centre Chalkboard, Brookings Institution. Retrieved from https://www.brookings.edu/articles/why-do-americans-rate-their-local-public-schools-so-favorably/

Wickham, H. (2016). *Ggplot2 : elegant graphics for data analysis* (2nd ed.). Springer.

Wiliam, D. (2011). *Embedded formative assessment*. Solution Tree Press.

Willis, J., Spina, N., Blundell, C., Spooner-Lane, R., Gordon, D., Arnold, J., & Gallagher, J. (2021). *QCAA accreditation course evaluation: Final report*.

Working Party on Tertiary Entrance. (1987). *Tertiary entrance in Queensland: a review*. Minister's Joint Advisory Committee on Post-Secondary Education and the Board of Secondary School Studies.

09
Appendices

9.1 Appendix A - Areas for review from Terms of Reference

Table 9.1 Areas for review from Terms of Reference and their corresponding report sections

Area for review		Report section
Rationale and foundation concepts		
1.1	The accessibility and quality of information on the rationale for the introduction of the new system, its theoretical and methodological underpinnings, and its essential policies and procedures.	3.1
1.2	How the objectives of the ACER review have been met in the government policy position and the new system.	3.1
Development and transition processes		
2.1	The success of the development process, including the various assessment trials and other testing activities that informed the development of current processes and procedures, and the effectiveness of the QCAA's program and project management.	3.2
2.2	QCAA's stakeholder communication and change management strategies to ensure that teachers and schools understand the new system and are appropriately skilled to exercise their responsibilities in their school context.	3.2
2.3	The organisational change process within the QCAA and whether appropriate systems are in place to ensure its internal workforce continues to develop and attract new expertise, embrace new ideas and ways of working, and become an employer of choice for teachers and other professionals.	3.2
Curriculum		
3.1	The QCAA's effectiveness in developing new syllabuses and resources and providing other support for curriculum development and delivery in schools. This should include consideration of the QCAA's syllabus development, review and revision processes and activities, and whether the new syllabuses meet the objectives set in the development phase (e.g., provide students with the opportunity to develop the 21st century skills, reflect best practice in teaching and learning, and are current and responsive to advances in knowledge, government initiatives and stakeholder needs).	4.1
New assessment processes		
4.1	Whether the key elements of the new assessment system are fit-for-purpose and their theoretical basis strengthened through implementation. The focus should be on the three main assessment processes of Endorsement, Confirmation, and External Assessment. Factors to consider are the interrelationships between the three processes, educational benefits to students, success in improving validity and reliability of assessments, value for money, and return on effort invested in their delivery.[1]	5.1
4.2	QCAA's strategy for the recruitment and retention of casual assessors to ensure there is an available and dependable workforce for each assessment process and activity.	5.1

[1] QCAA and UoM agreed value for money and return on investment could not be substantially addressed in this engagement. QCAA should consider an evaluation of these elements as part of its long-term program monitoring.

Supporting policies, procedures, and systems		
5.1	Whether the policies, procedures and supporting resources that comprise the QCE system are fit-for-purpose. This should include an assessment of whether the qualification is contemporary and supports students to follow their chosen pathways.	4.2
5.2	The documentation of processes and procedures to ensure effective functioning of the system and accountability to its stakeholders, including the QCAA staff and school personnel who play essential support roles (e.g., principals delegates, External Assessment coordinators, assessors).	5.2
*5.3**	*The redevelopment of QCAA's ICT systems to support the new QCE system. This should include an assessment of the reliability, accessibility and usability of the new systems, processes for evaluation or ongoing review of applications, and compliance with Queensland Government policies and guidelines.*	*5.2*
5.4	The suitability of access arrangements and reasonable adjustments to meet the needs of students whose access to education is affected by disability, impairment, medical condition, or other circumstances.	5.2
5.5	Whether the processes and systems for calculating and reporting subject results are well-defined, efficient, and transparent.	5.2
Ongoing monitoring and review		
6.1	Ongoing governance arrangements to ensure the system is monitored, reviewed, and continues to evolve to meet changes in the wider educational environment.	6.1
6.2	The use of data and other information to monitor the system and make improvements.	6.1
6.3	Availability and accuracy of data and other information to support schools, schooling sectors, QTAC and other stakeholders; and to ensure transparency and public accountability.	6.1
System impact		
*7.1**	*How well the system is understood by stakeholders and how it impacts on their day-to-day activities.*	*6.2*
7.2	The level of public confidence in senior assessment and certification since the new system commenced.	6.2
7.3	Whether the system maintains academic integrity and supports students to act responsibly and ethically.	6.2
7.4	The result of increasing subject-based and content experts at the QCAA to provide teachers with advice on teaching, learning and assessment.	6.2
7.5	Whether the new accreditation, training and professional development opportunities have led an improvement in teachers' expertise, especially in relation to their knowledge of assessment practice and capacity to make high stakes judgments.	6.2
*7.6**	*The impact of the reduction in summative assessments in school subjects, including its contribution to student well-being and whether it has increased teaching time.*	*6.2*
*7.7**	*The flexibility of the system to cater for the individual needs of students and varying school contexts (e.g., the delivery of alternative sequences and senior external examinations in some subjects; literacy and numeracy requirements, development of an international QCE).*	*6.2*
7.8	The fairness and accuracy of subject results.	6.2

* Areas for review 5.3, 7.1, 7.6 and 7.7 were excluded by agreement between QCAA and UoM on the grounds that there was insufficient disciplinary expertise in the UoM team to rubustly evaluate 5.3, and insufficient data were collected to support valid inferences about 7.1, 7.6 and 7.7.

9.2 Appendix B - Psychometric analyses

9.2.1 Rasch calibration, DIF and inter-rater analyses

To examine the psychometric properties of the QCAA assessments, nine general syllabus subjects were selected out of the 48 on offer, in collaboration with the QCAA. The aim was to have subject representation from each of the eight learning areas, as well as subjects that have diverse approaches to their Internal Assessments (IAs) and External Assessments (EAs), although it is acknowledged that these may not be representative of all the 48 subjects. The selected subjects were:

- Biology
- Business
- Design
- English
- Japanese
- Mathematical Methods
- Modern History
- Physical Education
- Visual Art

The item-level data (i.e., the criterion-level scores for the IAs and question-level scores for the EAs) collected in 2021, 2022 and 2023 were analysed for each subject. The data for 2020 was not included as schools only conducted two IAs due to the COVID-19 lockdowns. In addition, the Biology, Design, Japanese, Modern History and Physical Education subjects included data for alternative sequence (AS) students, which results from a flexible arrangement offered to schools by QCAA where they may teach the content of the syllabus in a different sequence from that stipulated in the general syllabus. This enables them to offer subjects by way of combined Year 11 and 12 classes that would otherwise have too low candidatures to be viable. The AS has common IAs relative to the general syllabus for all five subjects; however, the EA is unique relative to the general syllabus EA. Thus, for these subjects, the summary statistics presented below include the analyses for data from both the general syllabus and AS except for (a) evaluating differential item functioning (DIF) by background characteristics, as the sample sizes of the AS data were not sufficient for such analyses, and (b) the inter-rater agreement analyses, as we were not provided with this data for the AS students. The Japanese AS data for Japanese was not included in any analyses due to an insufficient sample size (n = 67) for robust psychometric analysis.

Figure 9.1 below illustrates the overall process of the psychometric analysis. Double-marked subjects are noted in step 1 with (2×Marked).

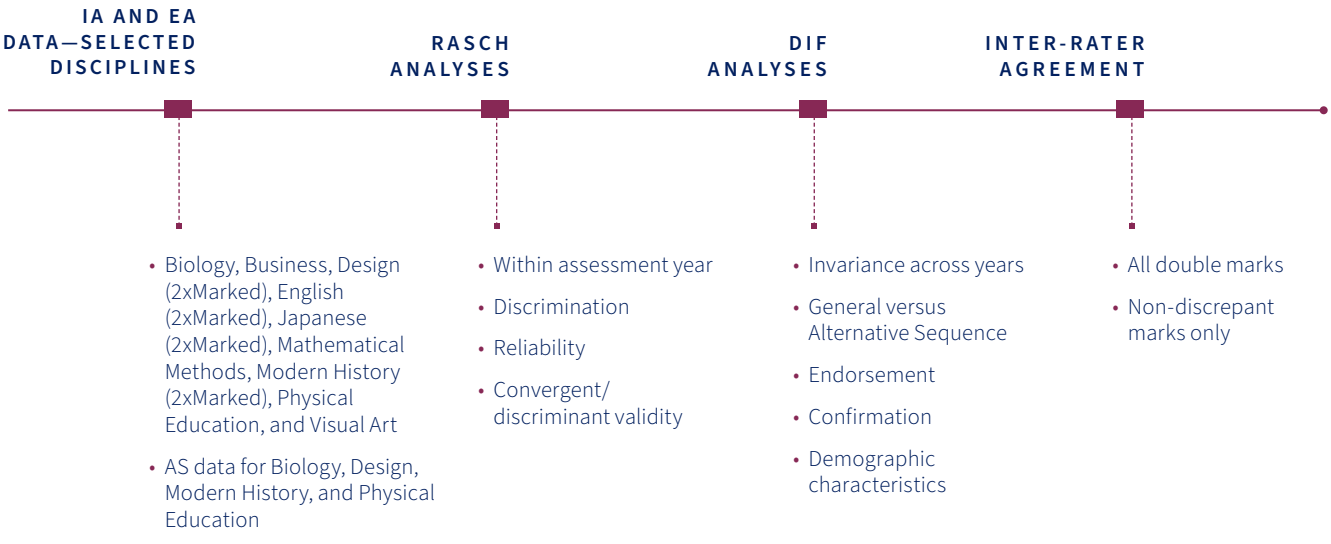


Figure 9.1 Psychometric data and analysis steps

Rasch analyses were conducted for each subject to calibrate the IA criteria and EA items by testing years. The main purpose of these analyses was to examine the psychometric functioning of the criteria and/or questions as well as the score reliabilities for the assessments. The correlations between the question- or criterion-level scores and total scores for the entire IA or EA (i.e., the item-total correlations) were calculated to evaluate discrimination values and identify how well the criteria and questions had performed in differentiating between the lower- and higher-achieving students. The criterion for low versus acceptable discrimination was set at 0.2 and the criterion for high discrimination was greater than 0.5. The correlations between students' scores on the IAs and EAs were then examined for each subject to evaluate the convergent/divergent validity between them. In addition, for the IAs Differential Item Functioning (DIF) analyses were conducted for each subject to examine whether all the IA criteria performed invariantly across the three calendar years (2021 to 2023) given the use of common instrument-specific marking guides (ISMGs). The same analyses were then conducted for the AS data across the five subjects. Further DIF analyses were conducted to examine whether the IA functioned invariantly across students taught the general syllabus sequence and the AS sequence.

In the next step, the effects of the Endorsement and Confirmation processes on the performances of the IA criteria were examined. Firstly, DIF analyses were conducted to examine whether all the IA criteria performed similarly between the assessments that were endorsed the first time versus those that were endorsed the second or third times. Secondly, DIF analyses were performed to examine whether all the IA criteria functioned invariantly before and after the Confirmation process. Additionally, fixed-parameter Rasch calibrations were conducted to equate the student ability estimates – which represent their overall scores on the IAs – obtained from both the provisional data and the confirmed data to examine how the Confirmation process impacted the students' ability estimates.

The third analysis involved examining whether the IA criteria and EA items displayed any biases in terms of different background characteristics of the students for the general syllabus data. Specifically, DIF analyses were conducted to examine whether the IAs and EAs functioned invariantly based on the following demographic variables:

- Gender – female and male students (treated as binary due to sample size requirements for DIF analyses)
- ATSI status – neither Aboriginal nor Torres Strait Islander origin or Aboriginal but not Torres Strait Islander origin/ Torres Strait Islander but not Aboriginal origin/both Aboriginal and Torres Strait Islander origin (treated as binary due to sample size requirements for DIF analyses)
- Home language – English and non-English (treated as binary due to sample size requirements for DIF analyses)
- School sector – government, Catholic and independent
- Local government area (LGA) – cities, regions and shires
- access arrangements and reasonable adjustments (AARA) status – AARA and non-AARA

Finally, for each of the four subjects with multiple raters for the EA items (Design, English, Japanese & Modern History), Fleiss's kappa statistics were calculated to examine their inter-rater agreement (IRA). Given that most of the criteria were polytomously scored by the markers, the quadratic weighted kappa statistics were calculated, as these factor in the size of the discrepancy in the assigned score and allocate less weight to smaller discrepancies. These analyses were separately conducted for each criterion for two different datasets for each subject in 2023. The first dataset included all students, including the cases where a third marker was required to moderate the two sets of scores (the third markers' scores were not included in the inter-rater agreement calculations, as they were not blinded judgements); the second dataset was the subset of data where the third marker was not required, as the scores across the two raters were the same or were not sufficiently discrepant according to the QCAA tolerance criterion for the subject. The results of the former dataset's analyses provide an indication of the overall inter-rater agreement, and the results of the latter give us an indication of the inter-rater agreement when there is no third marker intervention, thus providing insight into the level of inter-rater reliability (IRR) that is implicitly deemed tolerable by QCAA for the different subjects.

The above analyses were performed using the R programming language (R Core Team, 2024). For the Rasch analyses, the dichotomous Rasch (1960) model and the partial credit model (PCM, Masters (1982) were used to analyse the dichotomous and polytomous items respectively. The Rasch DIF model was used to run DIF analyses, which is an application of the many-facets Rasch model in which DIF is indicated by item–group factor interaction (Paek & Wilson, 2011). Based on Paek and Wilson's criteria, the items were categorised into four DIF categories between pairs of groups for each background factor: no or trivial DIF, small DIF, moderate DIF and large DIF. For the sake of these analyses, only cases of moderate and large DIF were considered as substantial DIF for an item, given that trivial and small instances of DIF typically have little to no influence at the whole assessment level. All the Rasch and DIF calibrations were performed using the R package TAM (Robitzsch et al., 2025). The inter-rater agreement analyses were conducted using the irrCAC package (Gwet, 2019).

9.2.2 Summary of the results

Table 9.2 summarises the analysis results, with the details of each described in the following sections. The table includes a brief description of the findings for:

- Any issues with Rasch model fit. It is not uncommon to have a few items with poor or slightly poor fit in any assessment. Many items showing misfit would suggest a careful review of the items, development process and potentially some cognitive labs, using think-aloud protocols with students to determine where the misunderstandings are occurring.
- The level of item discrimination. Item discrimination is a desirable characteristic, as it indicates the degree to which items separate the responses of high- and low-performing students on the item.
- The score reliability over time as calculated using the weighted likelihood estimates (WLE) of student ability. Generally, acceptable reliability for high stakes assessments is 0.80 or higher.
- The alignment between the mean difficulty of the assessment and the mean estimate of student ability. In the 'Difficulty match to student ability' column, the symbol ≈ indicates that the assessment difficulty and student cohort ability are closely matched; < or > means that the assessment difficulty is slightly higher or lower than the cohort ability level; << or >> means that the assessment difficulty is significantly higher or lower than the cohort ability level. To maximise the desired measurement properties of the assessment, these values should be closely aligned. If assessments are much higher than the mean student ability, students at the lower end of the ability scale may have difficulty accessing the assessment items; conversely, if the assessments are much lower than the mean student ability, high-performing students may not be sufficiently challenged and the assessment may not be effective in differentiating their performance.
- The correlation of the estimated student ability from the IAs and the EA. Moderate to high values indicate that the instruments are measuring the same construct and providing similar estimates of student ability.
- The impact of the Confirmation process, expressed as a percentage of scores not confirmed and then changed during the Confirmation of school provisional marking. If schools are accurately applying the ISMG with endorsed instruments that are aligned to the syllabus, the number of score changes is expected to be small. In addition, directionality of the changes (all/most scores increased or decreased) may indicate some level of marking bias in the provisional marks from the schools.
- Any DIF over time. If items or criteria show DIF over the years, it may indicate marking drift over time. DIF calculation is anchored in an assumption of overall equivalent performance, and the statistic detects deviations from this assumption.
- Any DIF in demographic groups. Items are assumed to function the same way for all students as part of the statistical modelling and as a societal expectation underpinning public confidence in the equity of an assessment. It is preferable that the focal group (usually the smaller or minority group in the analysis) and reference groups show statistically similar performance.
- For subjects with more than one marker evaluating the response, two sets of results are presented – inter-rater agreement (IRA) for all double-marked items and the IRA calculated on the subset of items whose marks were not deemed too far separated on the marking guide (discrepant).

The analysis results presented in Table 9.2 for the IAs are aggregated so that the three IAs in one year are treated as one test 'form' and then analysed over the three years of data provided by QCAA prior to analysis. This is done to increase the data set size and number of criteria in the analysis. Otherwise, the standard errors of measurement from the within-IA/within-year modelling would be so large as to make the analysis uninterpretable. Also note that these analysis results, except for the DIF and inter-rater agreement, are completed in the proficiency metric resulting from the Rasch analysis. Rasch analysis transforms raw scores (e.g., the number of correct answers on a test) into a Rasch ability measure, a latent trait usually represented as a theta (θ) value, with a fixed distance between adjacent points on the scale. This conversion allows for more meaningful interpretations of scores, especially when dealing with instruments where items have varying levels of difficulty, as is the case for the IAs and EAs.

Cells shaded in purple may be worth further investigation. Like any statistical measure, these flags have an associated error rate and should be read as indicative and considered and triangulated with other indicators that might suggest concern. Disaggregation of the data to a finer grain and analytic approaches focusing on specific concerns may provide illumination as to approaches to further improvement. As an example, Biology and Japanese IA1 each have a single criterion with a large scale (10 levels in Biology, 15 levels in Japanese). Having a single criterion is generally not considered to be best practice in assessment, as increasing the number of items or ratings will improve the reliability of the assessment. Rating scales of such a large size are also generally not recommended. There is an extensive literature in this area, and while more levels do increase reliability, this holds true only up to about seven points. Beyond seven levels there are diminishing returns on adding levels, and the cognitive load on markers is very high and may lead to inaccurate classifications.

Table 9.2 Calibration, DIF and inter-rater agreement analysis results by subject area and Internal/External Assessments

Subject area	Model fit issues	Item discrimination	Reliability (WLE) over time	Difficulty match to student ability	Correlation: IA/EA student ability	Confirmation impact	DIF: Time	DIF: Demographics	Inter-rater agreement
Biology: IA	1 criterion slightly poor fit	High	0.86–0.89	IA < SA	0.67–0.73	0.89% total 0.88% ↑, 0.01% ↓	None	One criterion LGA	
Biology: EA	1–3 items slightly poor fit	Mostly High	0.91–0.93	EA ≈ SA (2021, 2022) EA < SA (2023)		---	None	Few items Characteristics: Gender, ATSI, Home language, School sector, LGA	---
Business: IA	No issues	High	0.93	IA << SA	0.58–0.63	0.59% total 0.00% ↑, 0.59% ↓	None	None	
Business: EA	1 item slightly poor fit	High	0.75–0.84	EA ≈ SA		---	None	None	---
Design: IA	None	High	0.94	IA << SA	0.58–0.60	4.18% total 4.16% ↑, 0.02% ↓	None	None	
Design: EA	None	High	0.90	EA ≈ SA		---	None	None	0.40–0.60 (all) 0.63–0.71 (non-discrepant)
English: IA	None	High	0.92–0.93	IA << SA	0.61–0.62	0.12% total 0.01% ↑, 0.11% ↓	None	None	
English: EA	None	High	0.91–0.92	EA << SA		---	None	None	0.48–0.52 (KA, all) 0.30 (2 of 4 OD, all) 0.63–0.71 (KA, non-discrepant) 0.44–0.45 (2 of 4 OD, non-discrepant)
Japanese: IA	1 criterion slightly poor fit	High	0.88–0.89	IA << SA	0.74–0.76	1.93% total	None	None	
Japanese: EA	None	High	0.88	EA < SA		---	None	None	0.65–0.89 (all) 0.85–0.87 (non-discrepant)

Subject area	Model fit issues	Item discrimination	Reliability (WLE) over time	Difficulty match to student ability	Correlation: IA/EA student ability	Confirmation impact	DIF: Time	DIF: Demographics	Inter-rater agreement
Mathematical Methods: IA	None	High	0.83–0.85	IA << SA	0.79–0.80	0.11% total 0.01% ↑, 0.10%↓	None	None	
Mathematical Methods: EA	1–3 items slightly poor fit	Mostly High	0.93	EA < SA		---	None	Few items Characteristics: ATSI, Home language, School sector, LGA	---
Modern History: IA	1 item	High	0.93–0.94	IA << SA	0.55–0.64	0.28% total 0.03% ↑, 0.25% ↓	None	None	
Modern History: EA	1–3 items poor fit	Mostly High	0.82–0.86	EA << SA		---	None	Few items Characteristics: Gender, ATSI, Home language, School sector, LGA, AARA	0.56–0.90 (all) 0.64–1.0 (non-discrepant)
Physical Education: IA	2 items poor fit	High	0.92	IA << SA	0.63–0.64	0.85% total 0.01% ↑, 0.84% ↓	None	2 criteria Characteristic: Gender	---
Physical Education: EA	None	Mostly High	0.82–0.85	EA < SA		---	None	Few items Characteristics: Gender, ATSI, Home language, School sector, LGA	
Visual Art: IA	None	High	0.93–0.94	IA << SA		2.52% total 0.08% ↑, 2.44% ↓	None	None	
Visual Art: EA	1 item (2021) poor fit	High	0.92–0.93	EA << SA	0.52–0.54	---	None	1 item (2021) Characteristic: LGA	---

Biology

The Rasch analyses by year on the IAs indicated that the IA criteria measure a similar construct. There was only one criterion, IA1C1, that displayed slightly bad fit to the model; note that this is the only criterion on IA1. All the criteria had high item discrimination values, suggesting that they performed well in differentiating between lower- and higher-achieving students. The weighted likelihood estimate (WLE) score reliabilities ranged from .86 to .89 across the three testing years, indicating a good level of reliability. The average difficulty of the IAs was significantly lower than the average of the student ability estimates, and this was consistent across the three testing years. This suggests that students tended to receive high scores on IAs.

The results of the separate Rasch analyses by year for the EA indicated that most of the EA items had good fit statistics. Only 1 to 3 items (xea004.Q13.0042G2 in 2021, xea025.Q24a.0042G1 and xea031.Q26b.0042G1 in 2023) were found to have slightly bad fit per testing year. Most of the items had high item discrimination values, although a small number of items were found to have somewhat low discrimination values, which may be attributable to guessing response patterns. For the general syllabus students, the mean EA item difficulty was very close to the average student ability in 2021 and 2022 but was relatively lower than the mean student ability in 2023. For AS students, the average item difficulty of the EA was greater than the mean student ability in both 2021 and 2023 but similar in 2022. The WLE score reliabilities were high, ranging from .91 to .93 for the three years, indicating an excellent level of reliability.

The correlations between the IA and EA student ability estimates ranged from .67 to .73 for the three years, indicating a moderate level of correlation between the two.

None of the IA criteria were found to have substantial DIF across the testing years. In addition, no substantial DIF was observed in the IA criteria for the AS versus general syllabus data. Similarly, no substantial DIF was observed relevant to the Endorsement and Confirmation processes for the IA criteria. However, a very small percentage of students' scores were significantly affected by the changes of the marks due to the Confirmation process. There were about 0.89% (0.88% increased and 0.01% decreased) of students whose scores changed by more than half a standard deviation after the Confirmation process.

Almost no systematic DIF was found related to the student background variables. One IA criterion (IA3C4) was found to have DIF related to LGA. Only a few EA items were detected to have substantial DIF related to Gender, ATSI, Home language, School sector, or LGA in some years. All items across the IA and EA were found to perform invariantly between the AARA and non-AARA groups.

Business

The Rasch analyses by year indicated that none of the IA criteria had bad fit, suggesting that all IA criteria worked together to measure a similar construct. All the items performed well in discriminating lower- and higher-achieving students. The WLE reliabilities were high at around .93 for all three testing years, indicating an excellent level of reliability. The overall difficulty of the IA was significantly lower than the average of the student ability estimates for each of the three years, suggesting that students tended to receive high scores on IAs.

The Rasch analyses by year for the EA indicated that most of the external items had good fit to the model. Only one item in 2022 (xea009.Trends.0066G1) had somewhat bad fit and all items had high discrimination values. The WLE score reliabilities ranged from .75 to .84 for the three testing years, indicating a moderate level of reliability. The mean item difficulty was very close to or slightly below the average student ability estimates across the three testing years.

There was a moderate level of correlation between the IA and EA student ability estimates, ranging from .58 to .63 for the three years.

No IA criteria were found to have substantial DIF for the testing years. Moreover, no substantial DIF relevant to the Endorsement and Confirmation processes was observed for the IA criteria. However, a very small percentage of students' scores were significantly affected by the changes of the marks due to the Confirmation process. There were about 0.59% of students whose scores decreased but no students whose scores increased by more than half a standard deviation after the Confirmation process. No substantial DIF for either the IA criteria or the EA items was detected relating to Gender, ATSI status, Home language, School sector, LGA, or AARA status.

Design

The Rasch analyses by year indicated that all IA criteria measured a similar construct. All the criteria had high discrimination, suggesting that they performed well in differentiating between lower- and higher-achieving students. The WLE score reliabilities were .94 for all the years, indicating an excellent level of reliability. The average difficulty of the IA criteria was significantly lower than the average student ability estimates for each of the three years. This suggests that students tended to receive high scores on IAs.

The Rasch analyses by year indicated that all the EA items had good model fit statistics and high discrimination values. The WLE score reliabilities around .90 for all three years, indicating an excellent level of reliability. The average item difficulty was generally close to or slightly below the average student ability estimates for the three testing years.

There was a moderate level of correlation between the IA and EA student ability estimates, ranging from .58 to .60 for the three years.

No substantial DIF was observed for the IA criteria for the testing years. In addition, no substantial DIF relevant to AS versus general syllabus or for the Endorsement and Confirmation processes was detected for the IA criteria. However, a small percentage of students' scores were significantly affected by the changes of the marks due to the Confirmation process. There were about 4.18% (4.16% increased and 0.02% decreased) of students whose scores changed by more than half a standard deviation after the Confirmation process. None of the IA criteria

or EA items were found to have substantial DIF by Gender, ATSI status, Home language, School sector, LGA, or AARA status.

The IRA statistics for the general syllabus dataset with all students were moderate for all the criteria, ranging from .40 to .60. For the subset including only students who did not have a third marker the IRA statistics were moderate-to-good for all criteria, ranging from .63 to .71.

English

The Rasch analyses by year indicated that all IA criteria worked together to measure a similar construct. All the criteria had high discrimination values. The WLE score reliabilities ranged from .92 to .93 for all three years, indicating an excellent level of reliability. The average difficulty of the IA criteria was significantly lower than the average of the student ability estimates for each of the three years. This suggests that students tended to receive high scores on IAs.

The Rasch analyses by year indicated that all the EA items had good model fit statistics. All the items also had high discrimination values. The WLE score reliabilities, ranging from .91 to .92 for the three testing years, indicated an excellent level of reliability. The average item difficulty was significantly below the mean student ability estimate for each of the three testing years.

There was a moderate level of correlation between the IA and EA student ability estimates, ranging from .61 to .62 for the three years.

No substantial DIF was observed regarding the testing years for the IA criteria. Similarly, no substantial DIF was observed for the Endorsement and Confirmation processes for the IA criteria. However, a very small percentage of students' scores were significantly affected by the changes of the marks due to the Confirmation process. There was about 0.12% (0.11% decreased and 0.01% increased) of students whose scores changed by more than half a standard deviation after the Confirmation process. No substantial DIF was observed for the IA criteria or EA items by Gender, ATSI status, Home language, School sector, LGA, or AARA status.

The IRA statistics for the general syllabus data with all the students were moderate for all the knowledge application (KA) criteria, ranging from .48 to .52. For two of the four organisation and development (OD) criteria, the values of the IRA statistics were only fair (around .30). The values of the IRA statistics that were derived based on the subset of the data with the students who did not require a third marker were moderate-to-good for all the KA criteria and two of the OD criteria, with a range of .63 to .71. For the other two OD criteria, the IRA statistics were only fair (.44 and .45).

Japanese

The Rasch analyses by year indicated that most of the IA criteria had good model fit statistics. There was only one criterion, IA1C1, that had slightly bad fit in some testing years; as with Biology, this is the sole criterion on IA1. All the criteria had high discrimination values. The WLE score reliabilities ranged from .88 to .89 for the three testing years, indicating a good to excellent level of reliability. The average difficulty of the IA was significantly below the average student ability estimate for each of the three years. This suggests that students tended to receive high scores on IAs.

The Rasch analyses by year indicated that most EA items had good fit statistics. The discrimination values suggested all the items performed well in differentiating between lower- and higher-achieving students. The WLE score reliabilities were all around .88 for the three years, demonstrating a good level of reliability. The average item difficulty was slightly below the mean student ability estimate for each year.

There was a moderate to high level of correlation between the IA and EA student ability estimates, ranging from .74 to .76 for the three years.

No IA criteria were found to have substantial DIF relevant to testing year. No substantial DIF relevant to the Endorsement and Confirmation processes was found for the IA criteria. However, a small percentage of students' scores were significantly affected by the changes of the marks due to the Confirmation process. There were about 1.93% (1.88% decreased and 0.05% increased) of students whose scores changed by more than half a standard deviation after the Confirmation process. None of the IA criteria and EA items were found to have substantial DIF for Gender, ATSI status, Home language, School sector, LGA, or AARA status.

For EA scores for the general syllabus data with all the students, IRA statistics were moderate-to-good for all the criteria, ranging from .65 to .89. The IRA values for the data subset with only students who did not need a third marker were very good for all criteria, ranging from .85 to .87.

Mathematical Methods

The Rasch analyses by year indicated that all the IA criteria worked together to measure a similar construct. In addition, all the criteria had high discrimination values. The WLE score reliabilities ranged from .83 to .85 for all three years, indicating a good level of reliability. The average difficulty of the IA criteria was significantly below the average student ability estimates for each of the three years. This suggests that students tended to receive high scores on IAs.

The Rasch analyses by year indicated that most of the EA items had good model fit statistics. There were one to three items (xea013.Det.Var.0053G2, xea026.Tower.0053G2 and xea020.Internet.0053G2 in 2021; xea022.Q15.0053G2 in 2022; xea025.Q17.0053G2 and xea021.Q14c.0053G2 in 2023) in each year that had slightly bad fit statistics. Most of the items had high discrimination values. A few items were found to have slightly

low discrimination values in each year, which were consistent with guessing response patterns. The WLE score reliabilities were around .93 for the three years, indicating an excellent level of reliability. The average item difficulty was slightly below the mean student ability estimate.

There was a high correlation between the IA and EA student ability estimates, ranging from .79 to .80 for the three years.

No IA criteria were found to have substantial DIF by testing year. Similarly, no substantial DIF was observed in relation to the Endorsement and Confirmation processes for the IA criteria. However, a very small percentage of students' scores were significantly affected by the changes of the marks due to the Confirmation process. There was about 0.11% (0.10% decreased and 0.01% increased) of students whose scores changed by more than half a standard deviation after the Confirmation process. None of the IA criteria had substantial DIF relevant to Gender, ATSI, Home language, School sector, LGA, or AARA status. None of the EA items were found to have substantial DIF for Gender or AARA. Several EA items were found to have substantial DIF for to ATSI status, Home language, School sector and LGA in some testing years, but the small number would likely have had minimal impact on biasing overall exam performance.

Modern history

The separate Rasch analyses indicated that all the IA criteria except one, IA1C4, had good fit statistics. All the criteria had high discrimination values, suggesting that they performed well in differentiating lower- and higher-achieving students. The WLE score reliabilities ranged from .93 to .94 for all three years, indicating an excellent level of reliability. The average difficulty of the IA was significantly below the average student ability estimate for each of the three years. This suggest that students tended to receive high scores on IAs.

The results of the separate Rasch analyses by year for the EA indicated that most of the external items had good fit. There were 1 to 3 items in each year (xea004.Sim.0021G1 and xea005.Diff.0021G1 in 2021; xea002.Differ1.0021G1, xea001. Similarity.0021G1 and xea003.Differ2.0021G1 in 2022; xea023. Evaluate.0021G1 in 2023) that had relatively bad fit. Most of the items had high discrimination values, but a few were found to have slightly low discrimination. The WLE reliabilities ranged from .82 to .86 for the three years, indicating a good level of reliability. The average item difficulty was significantly below the mean student ability estimate for each of the three testing years. This suggests that the EA questions were relatively easier for the students.

There was a moderate level of correlation between the IA and EA student ability estimates, ranging from .55 to .64 for the three years.

None of the IA criteria were found to have substantial DIF relevant to the testing years. In addition, no substantial DIF relevant to AS versus general syllabus was detected for the IA criteria. Similarly, no substantial DIF was observed for the Endorsement and Confirmation processes for the IA criteria. However, a very small percentage of students' scores were significantly affected by the changes of the marks due to the Confirmation process. There was about 0.28% (0.25% decreased and 0.03% increased) of students whose scores changed by more than half a standard deviation after the Confirmation process. No IA criteria were found to have substantial DIF for Gender, ATSI, Home language, School sector, LGA, or AARA status. Only a few EA items were found to have substantial DIF relevant to Gender, ATSI status, Home language, School sector, LGA, or AARA status in some testing years, which would likely have had minimal impact on biasing overall exam performance.

The IRA statistics calculated for the data with all the students were moderate to excellent across the different criteria, with a range of .56 to .90. The values of the IRA statistics for the subset of this data with students who did not require a third marker, ranging from .64 to 1, were moderate to perfect across the criteria.

Physical Education

The Rasch analyses by year indicated that all IA criteria except two, IA1C2 and IA3C2, had good model fit statistics. All the criteria performed well in differentiating between lower- and higher-achieving students. The WLE score reliabilities were around .92 for all three years, indicating an excellent level of reliability. The average difficulty of the IA was significantly below the average student ability estimate for each of the three years. This suggests that students tended to receive high scores on IAs.

The Rasch analyses by year indicated that all EA items had good model fit statistics, suggesting that they worked well together to measure a similar construct. Most of the items had high discrimination values. A few items were found to have slightly low discrimination values in each year. The WLE score reliabilities ranged from .82 to .85 for the three years, indicating a good level of reliability. The average item difficulty was slightly below the mean student ability estimate for each of the three testing years.

There was a moderate correlation between the IA and EA student ability estimates, ranging from .63 to .64 for the three years.

No IA criterion exhibited substantial DIF by testing year. In addition, no substantial DIF for AS versus general syllabus was detected for the IA criteria. Similarly, no substantial DIF for the Endorsement and Confirmation processes was found for the IA criteria. However, a very small percentage of students' scores were significantly affected by the changes of the marks due to the Confirmation process. There was about 0.85% (0.84% decreased and 0.01% increased) of students whose scores changed by more than half a standard deviation after the Confirmation process. No IA criteria were found to have substantial DIF by ATSI status, Home language, School sector, LGA, or AARA status. Two of the IA criteria, IA1C2 and IA3C2, were found to have substantial DIF for Gender. Only a few EA items were found to have substantial DIF for Gender, ATSI status, Home language, School sector and LGA in some years. No EA items were found to have substantial DIF relevant to AARA status.

Visual Art

The Rasch analyses by year indicated that all the IA criteria worked together to measure a similar construct. All the criteria had high discrimination values, suggesting that they performed well in differentiating lower- and higher-achieving students. The WLE score reliabilities ranged from .93 to .94 for all three years, indicating an excellent level of reliability. The average difficulty of the IA was significantly below the average student ability estimate for each of the three years. This suggests that students tended to receive high scores on IAs.

The Rasch analyses by year indicated that all EA criteria had good fit statistics, except for one, xea005.Evaluate.0080G1, in 2021 with slightly bad fit. All the items had high discrimination values. The WLE score reliabilities ranged from .92 to .93 for the three years, indicating an excellent level of reliability. The average item difficulty was significantly below the mean student ability estimate for each of the three years. This suggests that the EA questions were relatively easier for students.

There was a moderate correlation between the IA and EA student ability estimates, ranging from .52 to .54 for the three years.

No substantial DIF relevant to the testing years was detected for the IA criteria. Similarly, no substantial DIF was found for the Endorsement and Confirmation processes for the IA criteria. However, a small percentage of students' scores were significantly affected by the changes of the marks due to the Confirmation process. There were about 2.52% (2.44% decreased and 0.08% increased) of the students whose scores changed by more than half a standard deviation after the Confirmation process. None of the IA criteria were found to have substantial DIF by Gender, ATSI status, Home language, School sector, LGA, or AARA status. In addition, no EA items exhibited substantial DIF by Gender, ATSI status, Home language, School sector and AARA status. Only one item, xea009.Terms.0080G1, was found to have substantial DIF relevant to LGA in 2021, but this would have minimal to no impact on the overall assessment performance.

9.2.3 Distributional analyses

The purpose of this set of analyses was to explore and evaluate changes in the IA and EA score distributions for the nine subjects (Biology, Business, Design, Japanese, English, Mathematical Methods, Modern History, Physical Education and Visual Arts) over time. Looking at equated score averages and distributions for both the IA and EA over time can provide various insights, including whether performance is improving since the QCE assessment reform due to greater familiarity among students, teachers, school and the system in general with the new qualification, or what is sometimes referred to as the sawtooth effect (Newton, 1997). Moreover, the relative improvement between IA and EA performance over time can provide evidence of possible scale drift: if the improvement in performance is markedly greater for the IA than the EA, it may indicate that the teachers' and schools' marking standards are drifting, with the same quality of work being given higher scores over time. Nonetheless, it is stressed that such findings are merely indicative and we cannot rule out other possible causes for improvements in performance over time for either the IA or EA: the data from 2021–2023 is also associated with greater time having passed since the disruptions caused by the COVID-19 pandemic, and the sawtooth effect may play out at different rates for IA and EA, and for individual subjects.

To conduct this exploratory analysis, it was necessary to equate the scores across the three years. For the IA, common instrument-specific marking guides (ISMGs) are used year on year, which provides a common item (i.e., the marking guide criteria) to equate (Kolen & Brennan, 2004). Consequently, given that the psychometric analyses did not show any DIF by testing year for the IA criteria, the scores were equated using a common-item non-equivalent groups (CINEG) design by running three separate Rasch analyses by year. For 2022 and 2023, the item difficulties (i.e., the threshold estimates between the ordered categories of the different ISMG criteria) were anchored to their 2021 values. Thus, the figures below for the IA distributions by year reflect the students' equated Rasch ability estimates across the three years. It should be noted that across the subjects, except for Japanese, the correlations between the freely estimated 2022/2023 IA difficulty estimates, and the 2021 calibrated estimates were ≥ .99, providing evidence for minimal impact of this anchoring to the 2021 values. For Japanese, the correlations were .85 and .87, but this was predominantly attributable to four thresholds estimates, two at the extreme low ability end and two at the extreme high ability end, which represented extreme categories with very few observations that were variable in number across the three years, thus creating large inter-year variability. Nonetheless, the results for Japanese should be interpreted with added caution given this greater relative variability in the difficulty estimates across the three years.

For the EA, the exams are unique from year to year and thus a different approach is required. While it would be possible to extend the CINEG equating applied to the IA across the three years to obtain equated estimates for the EA difficulties (as per the grade boundary setting analysis below), this would mean any marking drift in the IA scores across the three years would spuriously impact the relative EA difficulty estimates across the years. Given that the IA difficulties were constrained to be equal to their 2021 values in the above analysis, marking drift would be spuriously reflected in the modelled relative EA difficulties, making the latter years' estimates appear more difficult than they really were relative to 2021. Since we were specifically interested in independently comparing the distributional changes of the IA and the EA across the three years for the nine subjects, we adopted a kernel equating framework with a non-equivalent group with covariates (NEC) design to equate the 2022 and 2023 EA scores to their 2021 counterparts (Wiberg & Bränberg, 2015). This approach uses background factors

that are correlated with test performance, which in this case was a subset of the demographic variables outlined above for the DIF analyses (Gender, ATSI status, School Sector, LGA, Home language and AARA status), in place of common items to equate the non-equivalent groups' scores. It would be ideal to also include variable(s) related to the students' prior academic attainment on common assessments but that was not possible for the current analyses. The results of this equating are interpreted with some caution, particularly in terms of the more extreme scores. Specifically, the kernel equating used log-linear models for the pre-smoothing step and a Gaussian kernel for equating, where the latter is analogous to the equipercentile equating function but is applied to continuous (i.e., smoothed) rather than discrete scores (von Davier et al., 2004). A summary of the findings is provided below in Table 9.3.

All analyses were conducted using the R programming language (R Core Team, 2024), including the TAM (Robitzsch et al., 2025), ggplot2 (Wickham, 2016) and kequate (Andersson et al., 2013) packages. The resulting distributions are shown in violin plots. A violin plot is a statistical chart used to visualise the distribution of numeric data, particularly when comparing distributions between groups. It is like a box plot, showing the median value and inter-quartile range of the data. Violin plots also include a kernel density plot on each side, providing a more detailed view of the data's distribution. The width of the violin reflects the density of data points at different values, with wider sections indicating higher frequencies. In general, statistical distributions are expected to narrow at the extremes. Violin plots may indicate potential floor or ceiling effects if the lowest or highest estimated scores cluster at the bottom or top of the plot respectively.

9.2.4 Summary of the results

Table 9.3 Distributional analysis results by subject area and Internal/External Assessments

Subject area	IA: trend	IA: sawtooth effect	IA: ceiling or floor effect	EA: trend	EA: sawtooth effect	EA: ceiling or floor effect	IA & EA comparison: marking drift
Biology	Upward	No evidence	Ceiling, increasing over time	Stable	No evidence	Floor, small	Potential IA marking drift above the median
Business	Stable	No evidence	No evidence	Stable	Some evidence	No evidence	No evidence
Design	Stable	No evidence	No evidence	Stable	Some evidence	Ceiling, small but growing	Potential EA marking drift above the median
English	Stable	No evidence	Small ceiling emerging in 2023	Stable	Some evidence	Ceiling in 2022 only	Potential marking drift in 2022 only
Japanese	Stable	Some evidence	Small ceiling, but increasing over time	Stable	Some evidence	Floor, shrinking over time	Limited evidence
Mathematical Methods	Upward	Some evidence	Ceiling, increasing over time	Stable	Some evidence	No evidence	Potential IA marking drift
Modern History	Upward	Some evidence	Ceiling, increasing over time	Upward	Some evidence	Ceiling	Potential EA marking drift
Physical Education	Stable	No evidence	No evidence	Stable	No evidence	Ceiling in 2022 only	No evidence
Visual Arts	Upward	No evidence	Ceiling, increasing over time	Upward	Some evidence	Ceiling, growing 2022–23	No evidence

Biology

For Biology, there was a noticeable increase in IA performance across time, with an upward shift in the median and distribution each year and a growing ceiling effect whereby students achieved the maximum score for all three IAs. In contrast, the EA performance over time was quite stable other than a minor upward shift in the distribution from 2021 to 2022 and a slightly larger range of performance in 2023 relative to 2021 and 2022.

Combined, these observations are consistent with potential marking drift in the Biology IA results, particularly for students above the median. Even if drift in marking standards is not taking place, the current findings are concerning because, over time, the highest-performing students are not able to demonstrate the full extent of their attainment on the IAs.

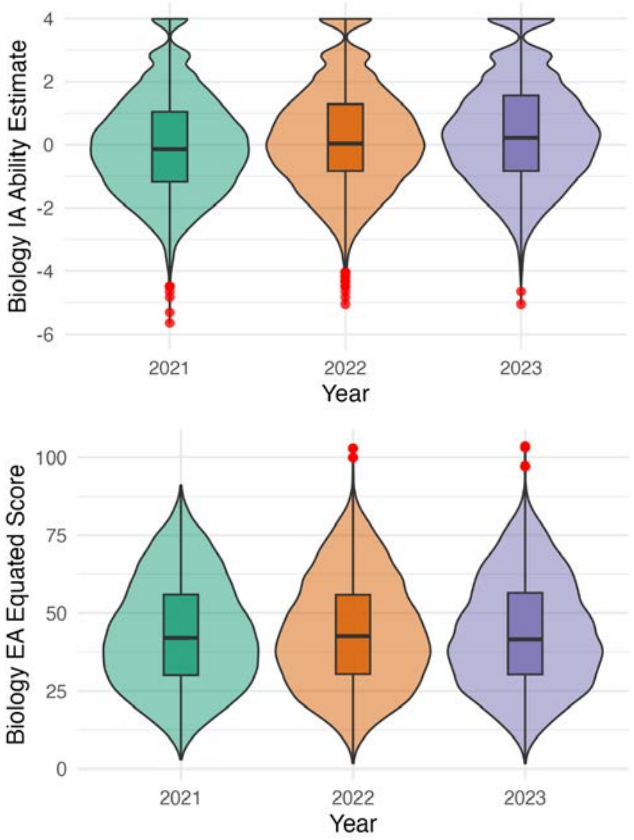


Figure 9.2 Violin plots for Biology showing the median and distribution of the IA ability estimates (top panel) and EA exam scores (bottom panel) across the three years

Business

For Business, performance on the IA has remained very stable over time other than a small shrinkage of the bottom end of the distribution in 2023. Similarly, performance on the EA has remained quite stable over time other than a very small increase in average performance in 2022 and 2023 relative to 2021 and an increasing stretching of the distribution above the median relative to 2021. Together, the findings support a very small sawtooth effect for the EA and do not show any evidence of marking drift for the IA.

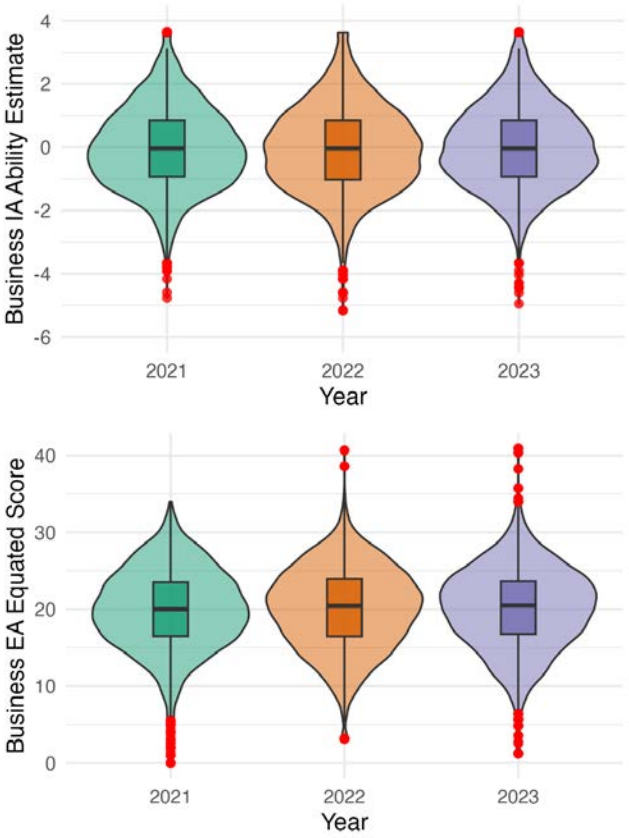


Figure 9.3 Violin plots for Business showing the median and distribution of the IA ability estimates (top panel) and EA exam scores (bottom panel) across the three years

Design

For Design, performance on the IA has remained very stable other than a slight average improvement in 2023 relative to the other two years. Similarly, performance on the EA has remained stable across the three years other than a very small increase in average performance in 2023, a small shrinkage of the distribution below the median performance in 2022 and an increasing stretching of the distribution above the median in 2022 and 2023 relative to 2021.

Together, these findings are indicative of a very small sawtooth effect for the EA, or a very small drift in marking standards for the EA over time. This could be of concern if the trend in the stretching of performance above the median continues leading to an increasingly pronounced ceiling effect where higher-performing students are limited in their capacity to demonstrate their attainment.

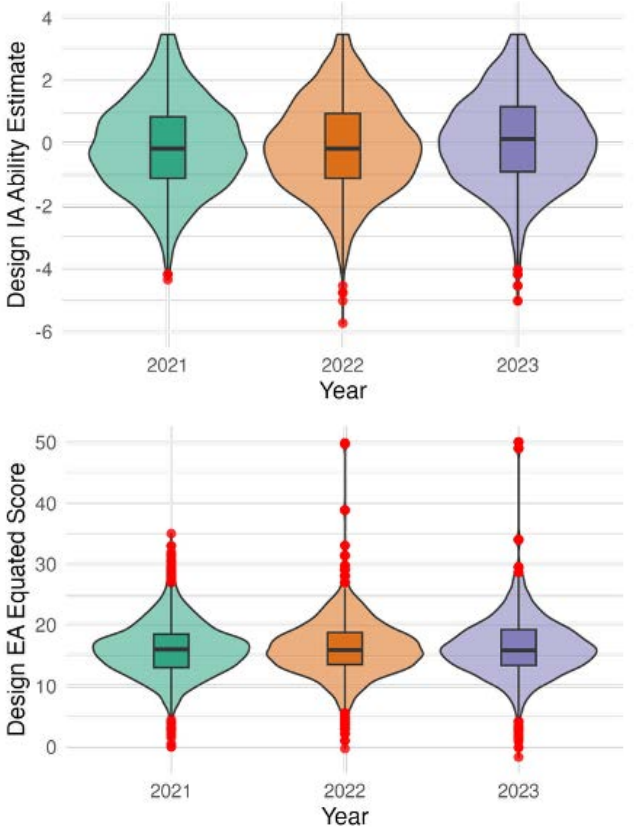


Figure 9.4 Violin plots for Design showing the median and distribution of the IA ability estimates (top panel) and EA exam scores (bottom panel) across the three years

English

For English, performance on the IA has remained quite stable other than a slight average improvement in 2022 and 2023 relative to the previous year/s, a very small but growing ceiling effect and a noticeable shrinkage of the distribution below the median in 2023 relative to the other two years. Similarly, performance on the EA has remained quite stable other than a small average improvement in 2022 relative to the other two years, including the emergence of a ceiling effect in that year, as well as a small shrinkage of the distribution below the median in 2023 relative to the other two years. Given the very similar average EA performance between 2021 and 2023, particularly for those performing above the median, there is evidence of a very small amount of marking drift in the IA, particularly in 2023 for students performing below the median. Moreover, there is evidence of a complex marker drift effect in the EA where there was a noticeable average increase in performance in 2022 that was not matched by increasing performance on the IA, but average performance then dropped in 2023 back to 2021 levels.

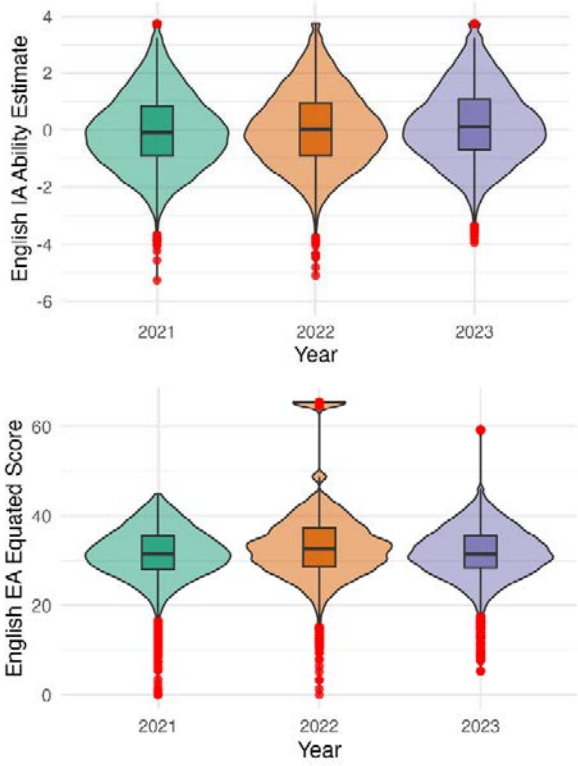


Figure 9.5 Violin plots for English showing the median and distribution of the IA ability estimates (top panel) and EA exam scores (bottom panel) across the three years

Japanese

For Japanese, performance on the IA has remained very stable over time other than a very small increase in average performance in 2023 relative to the other two years and the emergence of a small but growing ceiling effect across the three years. Performance on the EA shows a very similar trend, with a growing stretching of performance above the median across the three years and a small average increase in performance compared to the previous year/s along with a floor effect that appears to be shrinking over time as performance above the median stretches upward. Thus, there is evidence of a small sawtooth effect for both the IA and EA and limited evidence of marking standard drift across the years. The former is of concern if it continues to grow, as higher-performing students will be constrained in their capacity to demonstrate their full attainment.

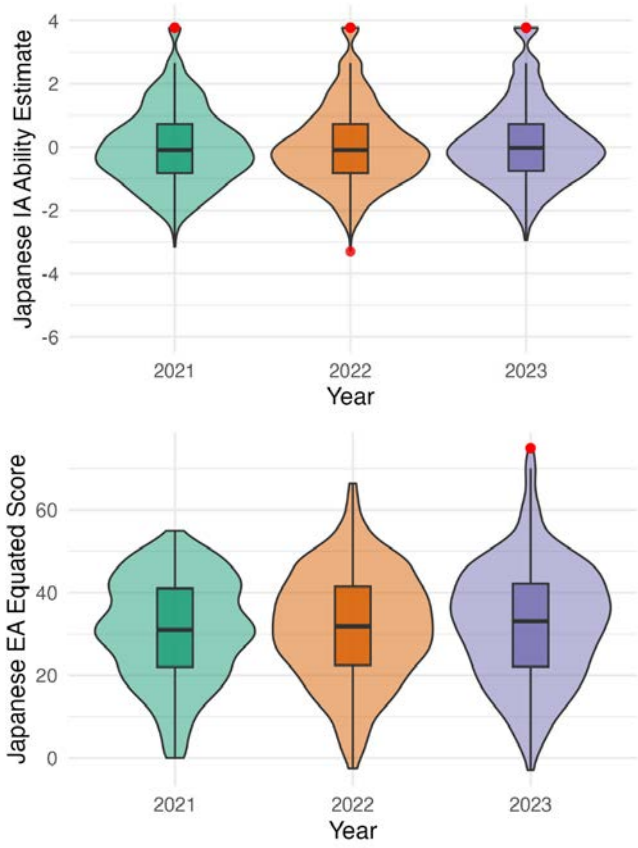


Figure 9.6 Violin plots for Japanese showing the median and distribution of the IA ability estimates (top panel) and EA exam scores (bottom panel) across the three years

Mathematical Methods

For Mathematical Methods, there has been a small increase in IA performance over time, with a small upward trend for both the median and distribution across the years and a growing ceiling effect. The EA performance has been quite stable but there has been a stretching of the distribution above the median over the three years and a very small increase in average performance in 2023 relative to the other two years. Combined, these findings indicate the presence of a sawtooth effect for both the IA and EA but given the magnitude of the increase in performance over time for the IA relative to the EA, there is also evidence of a degree of marking drift in the IA scores.

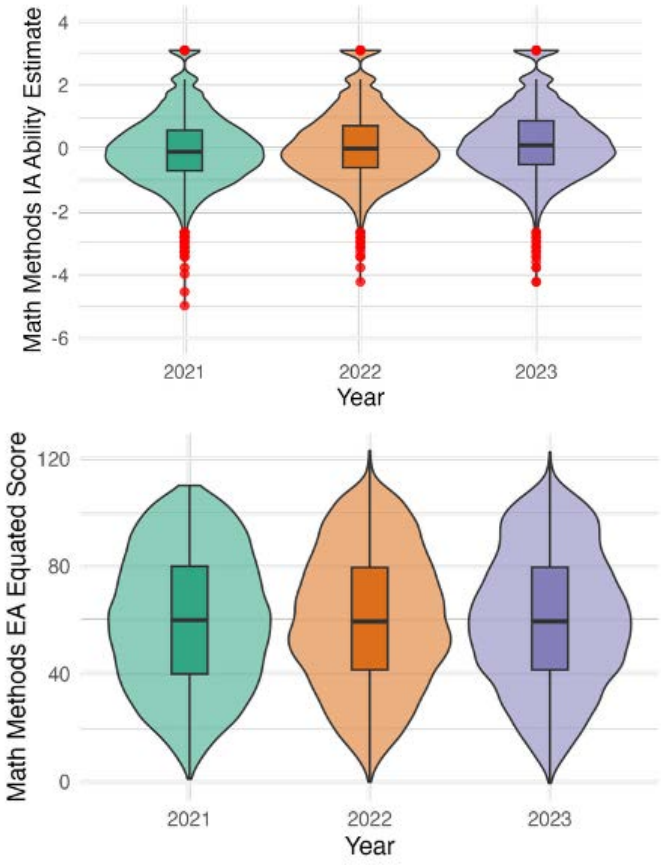


Figure 9.7 Violin plots for Mathematical Methods showing the median and distribution of the IA ability estimates (top panel) and EA exam scores (bottom panel) across the three years

Modern History

For Modern History, there has been a small upward trend in average performance and the distribution for the IA across the years, including a growing ceiling effect. This upward trend in the average and distribution, as well as the growing ceiling effect, is also present in the EA performance between 2022 and 2023. Combined, these findings indicate the presence of a small sawtooth effect and the growing ceiling effect is of concern given the impact on the highest-performing students' opportunity to demonstrate their full attainment.

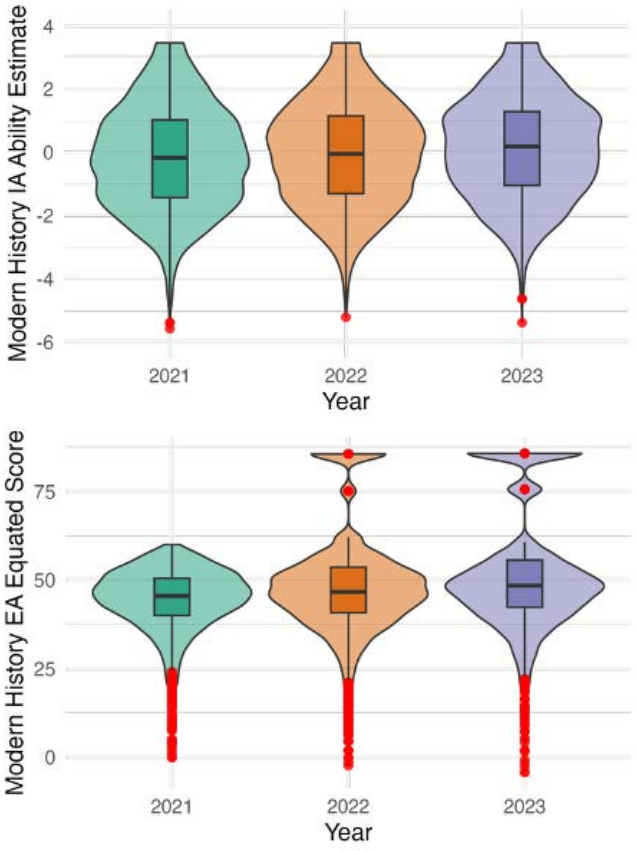


Figure 9.8 Violin plots for Modern History showing the median and distribution of the IA ability estimates (top panel) and EA exam scores (bottom panel) across the three years

Physical Education

For Physical Education, performance on the IA has remained very stable across the three years apart from a very small decrease in average performance in 2022 and a very small increase in average performance in 2023 relative to 2021. Similarly, performance on the EA has remained quite stable across the years apart from a very small increase in average performance in 2022 relative to the other two years. Given this stability, there is little evidence of any sawtooth effect or marking standards drift in Physical Education.

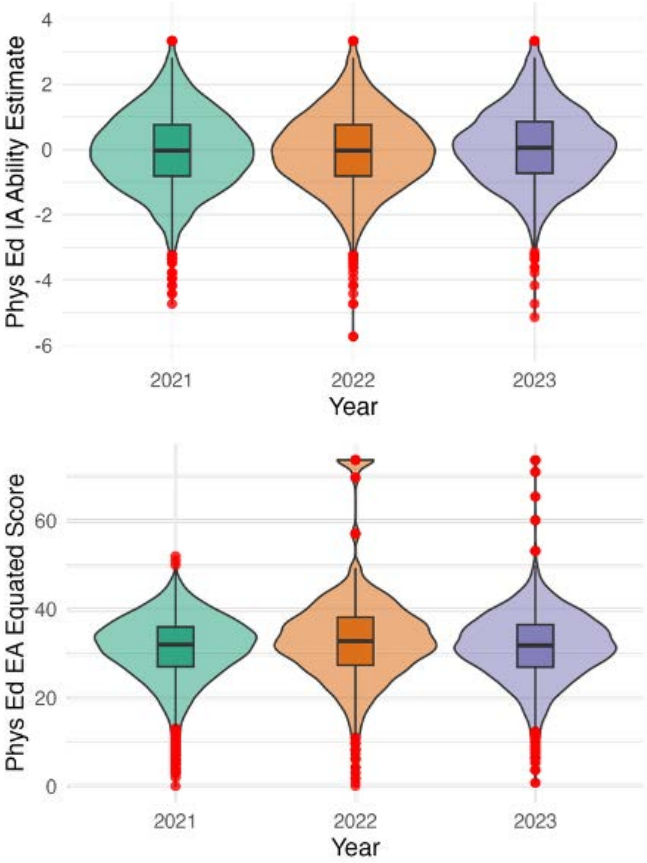


Figure 9.9 Violin plots for Physical Education showing the median and distribution of the IA ability estimates (top panel) and EA exam scores (bottom panel) across the three years

Visual Art

For Visual Art, there has been a small upward trend in IA performance across the years, including a growing ceiling effect. Similarly, the EA distribution performance has shown a small upward trend, with a small average increase in performance in 2023 relative to the other two years and a growing ceiling effect. Given the overlap in the distributional changes across the IA and EA, there is evidence of a small sawtooth effect but little evidence of marker drift across the years. The growing ceiling effect is of concern given the impact on the highest-performing students' opportunity to demonstrate their full attainment.

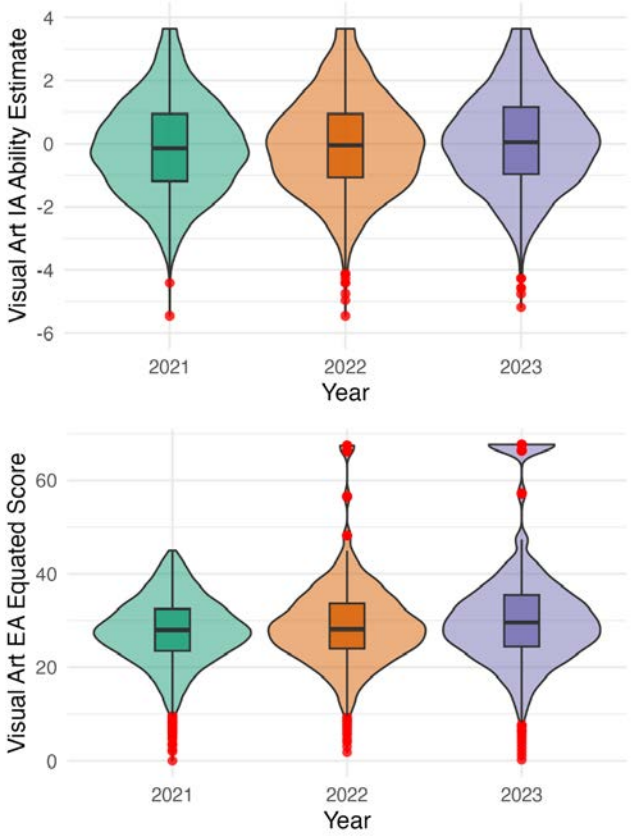


Figure 9.10 Violin plots for Visual Art showing the median and distribution of the IA ability estimates (top panel) and EA exam scores (bottom panel) across the three years

Conclusions

This exploratory analysis of the IA and EA score distributions across the nine subjects from 2021 to 2023 reveals nuanced patterns of performance and potential effects on marking standards over time. Across the subjects, several consistent trends emerge. Most notably, there appears to be a subtle but discernible sawtooth effect, characterised by gradual performance improvements over time as students, teachers, schools and the Queensland education system increasingly adapt to and become more familiar with the QCE assessment reform. However, the manifestation of this effect varies considerably between subjects and between IA and EA performances, with the EAs demonstrating more pronounced variability compared to the IAs.

Specifically, Biology showed a marked upward trend in IA performance relative to EA performance, with a growing ceiling effect that potentially limits the demonstration of high-achieving students' full attainment and is indicative of potential grade inflation. Other subjects, including Mathematical Methods and English, also exhibited greater upward trends in IA performance relative to EA performance, which could also indicate potential grade inflation. Again, it is stressed that these findings are merely indicative and there may be other causes for this increase in IA performance relative to EA performance, including the sawtooth effect having a differential impact between the IA and EA for certain subjects – although this seems unlikely for Biology and Mathematical Methods, as the psychometric analyses demonstrated a high correlation between the IA and EA for these subjects.

The analysis raises important considerations about assessment standards and their potential drift over time. While no dramatic systemic changes were observed, the consistent upward trends in several subjects' IA performances warrant careful monitoring. The growing ceiling effects, particularly in Biology and Mathematical Methods IA results, suggest that the current assessment instruments may not be fully discriminating among the highest-performing students. Methodological limitations should also be acknowledged, including that the kernel equating approach, while robust, was constrained by the unavailability of prior academic attainment variables. This limitation necessitates a cautious interpretation of the EA performance findings, especially regarding extreme scores. Furthermore, the ongoing impacts of the COVID-19 pandemic, particularly for the 2021 results, cannot be discounted.

Recommendations for the future include continued longitudinal monitoring of performance distributions, particularly in terms of the relative changes between IA and EA performance within a subject over time, refining some of the assessments to better differentiate high-performing students and investigating the underlying causal factors contributing to the observed performance trends.

9.2.5 Grade boundary setting analyses

The purpose of these analyses was to psychometrically evaluate the grade boundary setting for each of the nine subjects across the three years and, specifically, whether the expert judgement-based approach adopted by QCAA leads to grade boundary decisions that are consistent with the psychometric evidence. To conduct this evaluation, it was necessary to psychometrically equate the IA, EA and grade boundary logits across the three years. The IA and EA were equated using the CINEG approach described for the distributional analyses above. As discussed above, in the presence of marking drift for the IA this approach will lead to the EA difficulty estimates being spuriously easier for the earlier years compared to the next year/s; consequently, the EA difficulty distributions will be interpreted in this light for Biology and Mathematical Methods, and for 2023 English. To equate the grade boundary logits for 2022 and 2023 to their equivalent 2021 values, the robust Haebara linking method was used (He et al., 2015; Robitzsch, 2020). Specifically, for each of the subjects this involved running two separate Rasch analyses on the combined IA and EA data for 2022 and 2023, one with the IA difficulty estimates anchored to their 2021 values and the other with no such constraint. The linking analyses were then conducted to estimate the equating constants to transform the ability values of the unconstrained scale to their equivalent values on the anchored scale. These transformation constants for each year were then applied to the 2022 and 2023 grade boundary logit estimates. Given the assumptions of the modelling are reasonable, this approach illustrates variability in the grade boundaries over time. As part of this, it should be noted that across the subjects, the correlations between the freely estimated 2022/2023 IA and EA combined ability estimates and the equivalent estimates with the 2021 IA anchoring constraint were ≥ .999, demonstrating.

The results of these analyses are illustrated for each subject in the figures below. The top panel shows the distribution and median of the equated ability estimates from the combined IA and EA analyses for each of the three years. This enables us to evaluate the relative attainment of the different year cohorts, both in average terms and across the full ability scale, which should be reflected in the percentages of students attaining the higher or lower grades each year. However, it is noted that due to evidence of potential marking standards drift, these relative ability estimates should be interpreted with caution for Biology, Mathematical Methods and English.

The middle panel presents the distribution and median of the equated EA difficulty estimates, which illustrates whether the EA was easier or more difficult, both in average terms and across the full ability scale, across the three years. Ideally, assuming the EAs are functioning well, there should be minimal variation in the distribution of EA item difficulties from year to year, and any variability in these distributions should be reflected in the grade boundaries (i.e., if the EA was easier for the higher-performing

students in one year relative to the other two years, then the A/B grade boundary should be set to a more difficult location for that year). Again, these relative item difficulties should be interpreted with caution for Biology, Mathematical Methods and English. It is not necessary to evaluate the relative IA difficulties, as they were psychometrically constrained to be equal across the three years as described above.

The top two panels combined additionally provide information about how well the EA was targeted to the students' abilities across the years; ideally, there should be a large degree of overlap between the ability and difficulty distributions each year.

Finally, the bottom panel presents the equated grade boundary cutoffs across the years, including their 95% confidence intervals, which are based on the standard errors for the ability estimates that the grade boundary values represent. It is noted that the D/E boundary is set at an extreme value – i.e., at the extreme low end of performance – for each of the subjects across the three years, which means it has a large amount of equating error; comparisons of these equated boundaries across the years should therefore be avoided.

The following discussion of the grade boundary setting decisions across the years by subject considering the psychometrically equated scales is based on the actual percentages of grades achieved by year and subject (presented below in Table 9.4). This data was taken from the publicly available subject reports that are published annually by the QCAA for each of the QCE subjects. In this discussion, for each subject we consider: 1) changes in the ability distribution over time, including the overall reliability of these estimates; 2) the changes in the item difficulty distribution over time, including the overlap between the item difficulty and person ability distributions; and 3) the grade boundary decisions by year in light of differences in EA difficulty, as well as their relation to both the ability and difficulty distributions.

Biology

The equated person distributions across the three years for Biology show a steady increase in performance; the distribution was skewed to the lower end of performance in 2021, since when there has been an average improvement making the distribution more symmetrical and increasingly centred just above the 0-logit mark in 2022 and 2023, but with a growing tail toward the higher end of performance between 2022 and 2023. These findings are consistent with the suspected marker drift in the IAs, and so it is not possible to determine how much of this growth is an accurate reflection of the cohorts' relative attainment. Nonetheless, the ability estimates based on both the IA and EA showed an excellent level of reliability (WLE reliability = 0.95) for each of the three years.

With respect to the item difficulty distributions, the finding that the EA was noticeably easier on average in 2021 is consistent with the suspected marker drift, so again it is not possible to determine the extent to which this reflects a real difference in relative difficulty. Interestingly, this inflation is also suspected to have continued between 2022 and 2023, yet the 2022 EA is still observed to be slightly more difficult on average, suggesting that it may have been even more difficult relative to the 2023 EA. In terms of the overlap with the ability distribution, both the 2021 and 2022 EAs show good targeting to the ability levels of the candidates. The 2023 EA also shows reasonably good targeting, but the distribution of difficulty is bimodal with a noticeable increase in moderately easier (between −2 & −1 logits) items relative to the 2022 EA.

Table 9.4 Percentages of achieved grades by year and grade for the nine subjects

Subject area	Year	Grade percentage				
		A	B	C	D	E
Biology	2021	26.5%	46.7%	25.8%	0.9%	0.0%
	2022	30.0%	46.5%	22.4%	1.0%	0.0%
	2023	32.9%	44.8%	21.5%	0.8%	0.0%
Business	2021	17.4%	37.8%	39.2%	5.5%	0.1%
	2022	16.4%	36.0%	41.0%	6.4%	0.2%
	2023	16.7%	34.9%	43.9%	4.5%	0.1%
Design	2021	15.4%	38.5%	35.2%	10.6%	0.2%
	2022	15.9%	35.8%	38.3%	9.7%	0.3%
	2023	17.7%	35.4%	39.0%	7.6%	0.3%
English	2021	20.0%	41.9%	35.8%	2.3%	0.0%
	2022	19.4%	43.3%	35.6%	1.7%	0.0%
	2023	20.6%	43.8%	34.4%	1.2%	0.0%
Japanese	2021	38.7%	38.9%	18.7%	3.5%	0.2%
	2022	37.3%	39.8%	19.2%	3.6%	0.1%
	2023	38.3%	39.4%	19.1%	3.1%	0.1%
Mathematical Methods	2021	25.9%	38.2%	30.7%	5.1%	0.1%
	2022	27.0%	39.6%	29.6%	3.7%	0.0%
	2023	29.1%	40.0%	27.0%	3.9%	0.1%
Modern History	2021	29.6%	37.1%	30.6%	2.7%	0.0%
	2022	28.5%	38.4%	30.6%	2.5%	0.0%
	2023	31.3%	36.0%	30.1%	2.5%	0.1%
Physical Education	2021	19.0%	36.9%	37.4%	6.5%	0.3%
	2022	17.9%	36.7%	39.1%	6.1%	0.2%
	2023	17.8%	38.2%	39.0%	4.9%	0.2%
Visual Art	2021	18.1%	42.5%	30.9%	8.5%	0.1%
	2022	21.3%	38.0%	32.7%	7.8%	0.1%
	2023	21.9%	36.1%	34.0%	7.9%	0.1%

For the grade boundary setting, it is noticeable that both the D/E and C/D boundaries are set so far to the low-performing end of the scale that there is almost no overlap with the ability distribution for each of the three years; this is consistent with Table 9.4, with ≤ 1% of the candidates achieving a D or an E in any year. Interestingly, despite candidates located around the C/D boundary having greater opportunity to demonstrate their attainment on the EA in 2022 compared to the other two years, the grade boundary was set slightly easier in 2022. Moreover, for all three years, it is possible to achieve a C grade having only successfully achieved a very small number of score points on the EA. The B/C grade boundary across the three years shows that the decisions accurately reflected the relative difficulties of the EAs around that location, with the boundary set most difficult in 2021 and least difficult in 2022. On the other hand, the A/B grade boundary setting shows the opposite relationship with relative EA difficulty in and around that boundary location. This may reflect an attempt by the expert boundary setters to mitigate some of the inflation stemming from the IA scores. Nonetheless, the very small shifts in the A/B grade boundary relative to the rightward shifts in the ability distributions over the three years are consistent with the substantial growth in candidates achieving an A grade, and the steady decrease in those achieving a B or C, across the three years. Given that at least part of these shifts in the ability distributions is suspected to be attributable to IA inflation, it may be argued that these shifts in the A/B boundary should have been even further to the higher-performing end of the scale.

Business

The equated person distributions across the three years for Business indicate a relatively stable pattern of performance, with only minor shifts in the median ability levels over time. The 2021 distribution shows a slight negative skew, whereas in 2022 and 2023 the distributions are more symmetrical and increasingly centred just above the 0-logit mark. These trends suggest a modest upward shift in average performance over the three years, but the magnitude of change is relatively small. Importantly, there is no evidence of marker drift in the IA scores for Business, meaning the changes in the ability distributions can be more confidently interpreted as reflecting genuine differences in cohort attainment. The ability estimates based on both the IA and EA showed strong reliability in each year (WLE reliability = 0.94–0.95), supporting confidence in the precision of these estimates.

Regarding the EA item difficulty distributions, the 2021 and 2022 EAs are approximately equivalent in average difficulty, although they differ in shape. The 2021 EA shows a broad and bimodal distribution, with a substantial cluster of easier items located between −2 and −4, regions of the scale where there are relatively few students. This indicates a suboptimal targeting of items, with a noticeable portion of the test providing limited information for most candidates. In contrast, the 2022 EA displays a narrower, more unimodal distribution, better centred around the bulk of the student ability distribution and offering improved targeting overall. The 2023 EA is clearly more difficult on average, with its distribution shifted to the right and containing a greater proportion of harder items, with many of the item difficulties extending beyond even the highest-performing candidates. This shift suggests that relative to 2021 and 2022, the 2023 EA posed a greater challenge to students; this is reflected in the median difficulty being noticeably shifted to the higher-performance end of the scale. Nevertheless, the targeting in 2023 remains reasonable, particularly for higher-performing students, although it would benefit from more item difficulties being targeted toward the centre of the ability distribution.

For the grade boundary setting decisions, the D/E and C/D boundaries for all three years are located at the extreme lower end of the ability scale, with minimal overlap with the person distributions. This is consistent with Table 9.4, with fewer than 7% of candidates in any year receiving a D or lower. Specifically, the C/D boundary was set more difficult in 2021 relative to the other two years, which is consistent with the relative difficulties of the EA around that location; however, this grade boundary should have been set lower in 2022 relative to 2023 given their relative difficulties, which is reflected in the higher percentage of candidates achieving a D in 2022 relative to the other years. The equated B/C grade boundary location has been quite stable but very slightly increased every year; this impacted the 2023 cohort – who sat a substantially more difficult EA – as reflected in Table 9.4, with a noticeably larger proportion of students achieving a C compared to a B in 2023 relative to the other two years. The A/B grade boundary exhibits a small but clear rightward shift in 2022 and 2023 relative to 2021, which is consistent with their relative EA difficulties around that location, but potentially lower for the 2021 cohort given the overall proportion of easier items relative to the other two years. Consequently, despite the equated ability estimates showing a small rightward shift in the distributions for 2022 and 2023 relative to 2021, a smaller proportion of students in both 2022 and 2023 achieved an A grade.

Design

The equated person distributions across the three years for Design reveal a high degree of consistency in cohort performance. The distributions for 2021, 2022 and 2023 are closely aligned, each showing a slightly negative skew with very little change in central tendency over time. The median ability levels remain stable across the three years, suggesting minimal shifts in overall student attainment. Importantly, there is no evidence of significant marking drift in the IA scores, and the consistency in the person distributions supports the conclusion that any differences in grade outcomes likely reflect changes in assessment difficulty or boundary decisions rather than changes in cohort ability. The WLE reliabilities for the combined IA and EA scores across the three years were high (0.95), supporting confidence in the precision of these estimates.

In terms of the EA item difficulty distributions, the 2021 EA is clearly the most difficult on average, with its distribution shifted notably to the right relative to the student ability distributions. It also shows a relatively flat and wide profile, indicating a broad range of item difficulties, but with a concentration of more demanding items that exceed the ability levels of most candidates. In contrast, the 2022 and 2023 EAs are very similar in both average difficulty and the shape of their distributions, although the 2022 EA was somewhat easier than its 2023 counterpart. Both show a bimodal distribution that is better aligned with the student ability distributions than 2021, although many of the item difficulties are still located at extreme high- and low-performing locations. Overall, this suggests that there is redundancy in both the bottom and top ends of the marking guides for this subject and they could be revised to better target the attainment levels of the cohorts.

Regarding the grade boundary setting decisions, for both the D/E and C/D grade boundaries the locations were set as more difficult in 2021 relative to both 2022 and 2023, which is not consistent with the relative difficulties of the EAs at those locations. Consequently, Table 9.4 shows that the proportion of students achieving a D or lower was highest in 2021. In contrast, both the B/C and A/B boundaries were set lower for 2021 compared to the other two years, and the A/B boundary was set lower in 2023 relative to 2022, which all accurately reflect the relative difficulties of the EAs at those locations. However, the B/C boundary being set higher in 2023 relative to 2022 is not consistent with their relative EA difficulty distributions; this is reflected in the higher proportion of students achieving a C rather than a B in 2023 compared to 2022. For the higher-performing students, the small average increase in performance in 2023 compared to the other two years appears to be accurately reflected in Table 9.4, with a higher proportion of students achieving an A grade.

English

The equated person distributions across the three years for English suggest a modest upward shift in cohort performance from 2021 through 2023. The 2021 distribution is slightly skewed to the lower end, while the 2022 and 2023 distributions are increasingly centred around a higher median, with the 2023 cohort showing the highest concentration of students above the 0-logit mark. These shifts suggest some genuine improvement in average student attainment over time, although, as noted above, there is evidence of potential marking standards drift affecting the IA scores in 2023. Consequently, caution is warranted when interpreting the increase in ability levels across the years. Nevertheless, the ability estimates based on the combined IA and EA scores showed high reliability across all three years (WLE reliability = 0.95), supporting the precision of the estimates.

In terms of the EA item difficulty distributions, the 2021 EA is slightly more difficult on average than those of 2022 and 2023, although the overall differences are relatively minor. All three years show a broadly similar bimodal distribution, with the most prominent mode targeted to the easy end relative to the ability distribution, and the more minor mode targeted to the difficult end. Overall, this suggests that there is redundancy particularly at the low-performing end but also to some degree at the top-performing end of the marking guides for this subject, and they could be revised to better target the attainment levels of the cohorts.

Regarding the grade boundary setting decisions, the D/E boundary was set at a very extreme low-performing scale location, which is outside the bounds of any of the ability distributions across the three years. The C/D, B/C and A/B boundaries were consistently set lower in 2023 relative to the other two years, which is inconsistent with the relative EA difficulty distributions, both on average and around those locations, and is concerning given the suspected marker drift in the IA for this cohort. Consequently, Table 9.4 shows that the 2023 group achieved the highest proportion of A and B grades despite almost perfect overlap in the 2022 equated ability distribution above the median performance. Moreover, the A/B grade boundary was set at a higher in 2022 relative to 2021, which is inconsistent with the relative EA difficulty distribution around that location; this is reflected in Table 9.4 with the 2022 cohort achieving the lowest proportion of A grades despite a noticeable right shift in the ability distribution relative to 2021.

Japanese

The equated ability distributions across the three years for Japanese show a remarkable level of consistency, with only very minimal shifts in mean performance across the three years. The distributions for 2021, 2022 and 2023 are tightly centred around the 0-logit mark, with very slight rightward shifts observed across the three years. There was no indication of marking drift in the IA scores, and the equated ability estimates demonstrated high reliability across all three years (WLE reliability = 0.95), supporting the precision of the estimates.

The EA item difficulty distributions display a modest but interpretable degree of variation. The 2021 EA is the easiest on average, with its distribution shifted furthest to the left but still quite well aligned with the ability distribution. The 2022 and 2023 EA difficulties are almost identical on average, but the 2022 distribution is more compact and very well targeted to the 2022 ability distribution, whereas the 2023 distribution is broader, with more item difficulties located at both the low- and high-performing ends of the scale, although the latter also target the greater prevalence of ability estimates at the high-performing end of the scale in 2023. Despite these differences, the variation in EA difficulty is relatively modest, and all three EAs appear reasonably well targeted to their respective cohorts.

Regarding the grade boundary setting decisions, the D/E boundary was set at a very extreme low-performing scale location, which is almost outside the bounds of any of the ability distributions across the three years. Moreover, for all three years it was possible to achieve a D grade having only successfully achieved a very small number of score points on the EA. The C/D boundary was set lower in 2022 relative to 2021, but this is consistent with the relative EA difficulty distributions around that location. On the other hand, the 2023 C/D boundary was set lower in 2023 compared to 2021, which is inconsistent with their relative distribution of EA difficulties around that scale location; this is reflected in Table 9.4, with the 2023 cohort achieving the highest proportion of C grades relative to D grades of the three years. The B/C and A/B grade boundary scale locations in 2022 and 2023 generally follow the relative distribution of EA difficulties around those scale locations. However, they are inconsistent for the 2023 cohort relative to 2021, so despite the 2023 group showing a slight increase in attainment as reflected in the equated ability estimates compared to 2021, fewer achieved an A grade. Nonetheless, given the overall stability in ability distributions and minimal variation in EA difficulty, the grade boundary decisions for Japanese appear broadly consistent with the psychometric evidence.

Mathematical Methods

The equated ability distributions across the three years for Mathematical Methods show a steady and modest upward shift in cohort performance from 2021 through 2023. The 2021 distribution is more negatively skewed, whereas the 2022 and 2023 distributions become increasingly symmetric and centred around slightly higher medians. The 2023 distribution has the highest proportion of students performing above the 0-logit mark. However, as noted earlier, the presence of suspected marking standards drift in the IA scores for Mathematical Methods means that these shifts in attainment should be interpreted with caution. Despite this, the equated ability estimates for all three years demonstrated high reliability (WLE reliability = 0.96), allowing for confidence in the precision of the estimates within each year cohort.

The EA item difficulty distributions show clear variation across the years. The 2021 EA is the least difficult on average, with a distribution that is noticeably to the left of the other two distributions and is more bimodal in nature. The 2022 EA is the most difficult on average with a difficulty distribution that is more compact. The 2023 EA average difficulty is in between the

other two years, and the distribution is even more compact and very slightly left skewed. All three years show strong targeting of the EA distribution to the ability distribution, and this is especially the case for the 2023 EA.

Regarding the grade boundary setting decisions, the D/E boundary was set at the very low end of the scale in all three years, just overlapping the extreme left tail of the student distributions. This aligns with Table 9.4, with very few students (≤ 0.1%) receiving an E grade in any year. For the C/D, B/C and A/B grade boundaries, despite the 2021 EA being noticeably more difficult than the 2021 EA, and the relative EA difficulty distributions at the grade boundary scale locations indicating it was easier or equivalent difficulty to achieve those grades in 2021, the 2022 grade boundaries were consistently set at higher locations. This may reflect an attempt by the expert boundary setters to mitigate some of the suspected inflation stemming from the IA scores. However, the 2023 grade boundaries relative to 2021 are consistent with the relative EA difficulty distributions around the boundary scale locations, and there is no evidence of trying to mitigate any potential inflation in the IA scores. Thus, the 2023 cohort achieved the highest percentage of A and B grades of the three years, as reflected in Table 9.4.

Modern History

The equated ability distributions across the three years for Modern History show a noticeable upward shift in cohort performance from 2021 to 2023. The 2021 distribution is centred just below the 0-logit mark, with the 2022 and 2023 distributions becoming increasingly right-shifted and slightly more dispersed. The 2023 cohort has the highest proportion of students performing above the average scale location. There was no evidence of marker drift in the IA scores for this subject, meaning that the differences in ability distributions likely reflect genuine variation in cohort attainment. The equated ability estimates for all three years demonstrated high reliability (WLE reliability = 0.95), allowing for confidence in the precision of the estimates.

The EA item difficulty distributions exhibit a clear pattern of change across years. The 2023 EA is by far the easiest on average, with its distribution noticeably left-shifted relative to both 2021 and 2022. In contrast, the 2021 and 2022 EAs are more similar in average difficulty and show a large degree of overlap. All three distributions show a degree of bimodality with item difficulties that extend to the lower-performing end of the scale beyond the ability distribution, but this is most marked for the 2023 EA distribution. The 2021 and 2022 EA item difficulty distributions show reasonable targeting to their respective ability distributions, whereas the 2023 EA distribution is reasonably targeted to lower- to and higher-performing students.

Regarding the grade boundary setting decisions, the D/E boundary was set at an extremely low-performing location in all three years, consistent with the small proportion of students receiving an E grade each year (0.0%–0.1% in Table 9.4). The C/D boundary was set lower in both 2022 and 2023, which is inconsistent with the relative item difficulty distributions around that scale location and particularly should have been higher in 2023. Nonetheless, the locations of this boundary across the three years barely overlap with the lowest-performing tail of the ability distributions and so there is very minimal impact on the relevant grade percentages. In contrast, the B/C grade boundary scale locations across the three years are consistent with the relative item difficulty distributions around that scale location. For the A/B grade boundary scale locations, despite almost perfect overlap in the 2021 and 2022 EA item difficulty distributions in that region of the scale, the 2022 boundary was set noticeably to the high-performing end relative to 2021, which is reflected in the drop in the percentage of students in the 2022 cohort achieving an A grade, as shown in Table 9.4, despite the right shift in their equated ability distribution. Similarly, the A/B boundary location was set even further to the higher-performing end of the scale for the 2023 cohort, but the magnitude of the right shift in their ability distribution meant that the proportion of students in this group achieving an A grade was greater than for the 2021 cohort.

Physical Education

The equated ability distributions across the three years for Physical Education show an upward shift in cohort performance from 2022 to 2023. The 2021 and 2022 distributions are nearly identical in central tendency, both centred just below the 0-logit mark and highly overlapping in shape. In contrast, the 2023 ability distribution is modestly right-shifted, indicating a small increase in overall student attainment. There is no evidence of marker drift in the IA scores for this subject, suggesting that any observed differences in performance likely reflect genuine variation in cohort ability. The equated ability estimates demonstrated high reliability across all three years (WLE reliability = 0.94), supporting confidence in their precision.

The EA item difficulty distributions show clear differences across the years. The 2021 EA is the least difficult on average, with its distribution noticeably left-shifted relative to 2022 and 2023. It also shows a bimodal shape, with item difficulties most targeted to the lower- and higher-performing locations of the scale. The 2022 EA is more difficult, with a broad distribution that is very well targeted to the ability distribution, although there is some redundancy in item difficulties that go beyond the attainment of the cohort. The 2023 EA is the most difficult on average, with a right-shifted distribution that is narrower than the other two, and it is slightly mistargeted to the higher performance end of the scale. Nonetheless, it still shows good overall targeting of the ability distribution.

Regarding the grade boundary setting decisions, the D/E boundary was set at a very low-performing location in all three years, with only minimal overlap with the ability distributions. This is consistent with the very small proportion of students receiving an E grade each year (≤ 0.3% in Table 9.4). The C/D boundary was appropriately higher in 2021, aligning with the easier EA, while the 2022 and 2023 boundaries were essentially equal, which is inconsistent with the slightly easier distribution

of the 2022 difficulties around this scale location compared to the 2023 distribution. For the B/C grade boundary, the locations are inconsistent with the distribution of the EA difficulties, which would have negatively impacted both the 2022 and 2023 cohorts, and particularly the latter, which is reflected in the higher relative percentage of C grades for these cohorts shown in Table 9.4. The A/B grade boundary has been set progressively higher across the three years, which is consistent with the relative distribution of the EA difficulties around that scale location; however, given the distributional overlap between the top end of the 2021 and 2023 equated ability estimates, the fact that a smaller proportion of students achieved an A grade in 2023 suggests that the boundary may have been set too high for the 2023 cohort.

Visual Art

The equated ability distributions across the three years for Visual Art are highly consistent in shape and central tendency, with each year being slightly right-shifted with respect to the previous year, indicating a small, steady increase in overall student attainment. There is no evidence of marker drift in the IA scores for Visual Art, so these differences likely reflect genuine variation in cohort performance. The equated ability estimates demonstrated high reliability in all three years (WLE reliability = 0.95), supporting confidence in their precision.

The EA item difficulty distributions are remarkably stable across the three years. All three show similar shapes and means, with the 2023 EA marginally more difficult on average than 2021 and 2022. Each distribution is broad and relatively flat, with slight bimodality and item difficulties extending to both the low- and high-performing ends of the scale. This broad coverage ensures that the EAs provide very good targeting across the full range of student performance in all years, although there is some redundancy at the extreme low end of the scale where there are relatively few students.

Regarding the grade boundary setting decisions, the D/E boundary was set at an extreme low end of the scale in all three years, barely overlapping the ability distributions. This is consistent with the very small percentage of students receiving an E grade across years (0.1% in Table 9.4). The C/D boundary was set lowest in 2023 and highest in 2021, which is consistent with their relative EA difficulty distributions, but was set too low 2022 relative to 2021 given the slightly easier EA difficulty distribution around this scale location in 2022. The B/C boundary has been consistently set higher across the three years, which is not consistent with the relative EA difficulty distributions and would have been most impactful for the 2023 cohort given the overall greater difficulty of the EA relative to the other two years. The A/B boundary has similarly been consistently set higher across the three years with an even greater difference to the 2021 boundary than the B/C boundary. This does not appear to be supported by differences in the relative EA difficulty distributions and so may be influenced by other factors, such as not wanting the percentage of A grades to increase too dramatically between years, or an attempt to mitigate perceived inflation or drift in the IA scores.

Biology

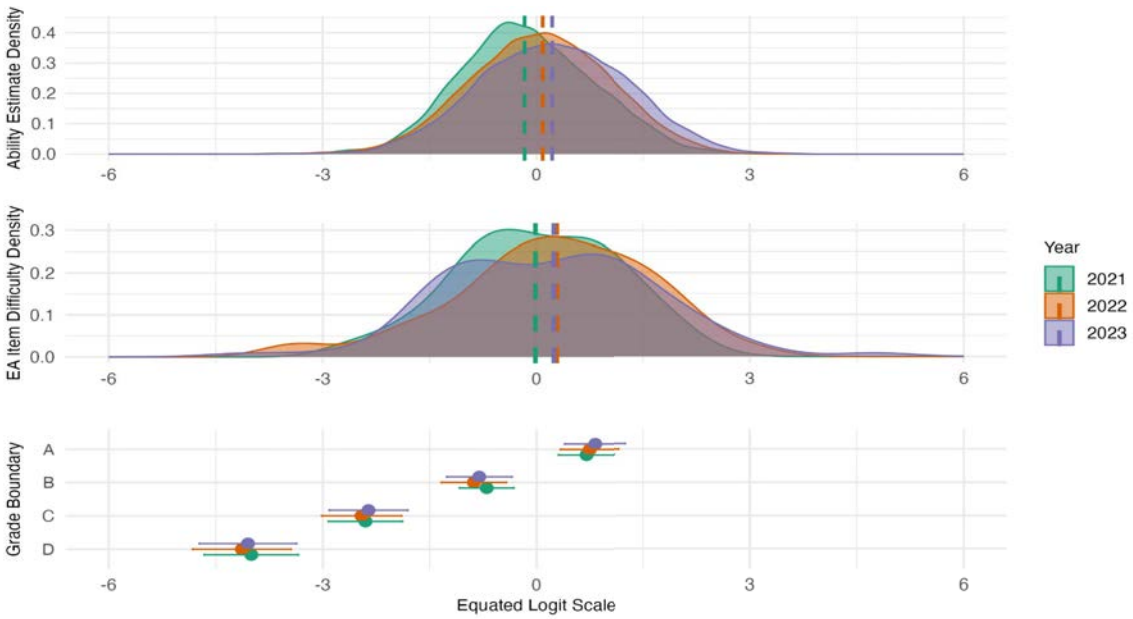


Figure 9.11 Plots for Biology showing the distribution and median of the equated ability estimates from the combined IA and EA Rasch analysis (top panel), the distribution and median of the equated EA item difficulty estimates (middle panel) and the equated grade boundary cutoffs with their 95% confidence intervals (bottom panel) across the three years

Business

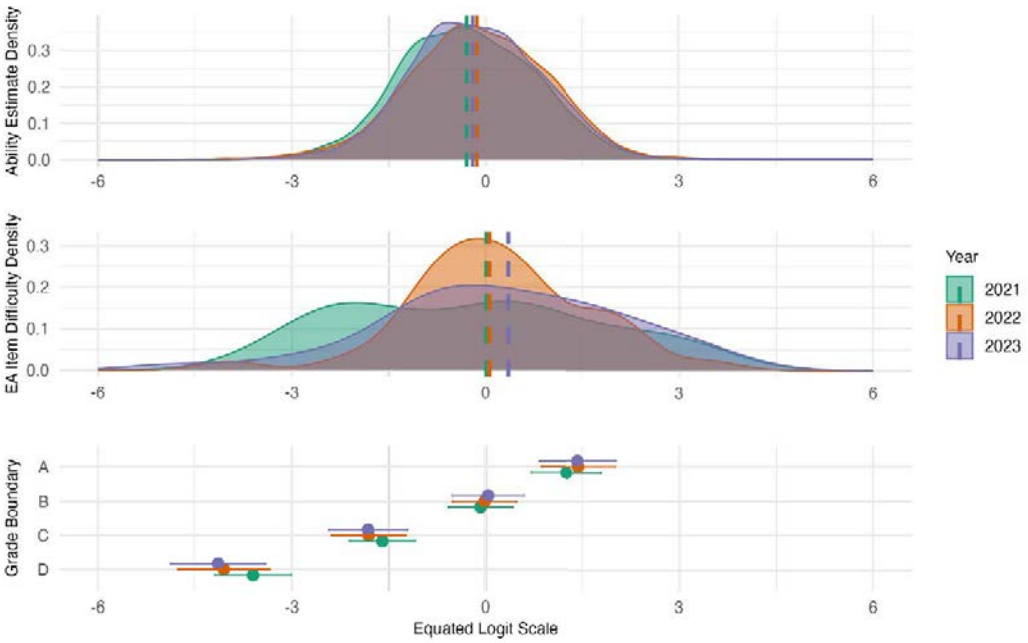


Figure 9.12 Plots for Business showing the distribution and median of the equated ability estimates from the combined IA and EA Rasch analysis (top panel), the distribution and median of the equated EA item difficulty estimates (middle panel) and the equated grade boundary cutoffs and their 95% confidence intervals (bottom panel) across the three years

Design

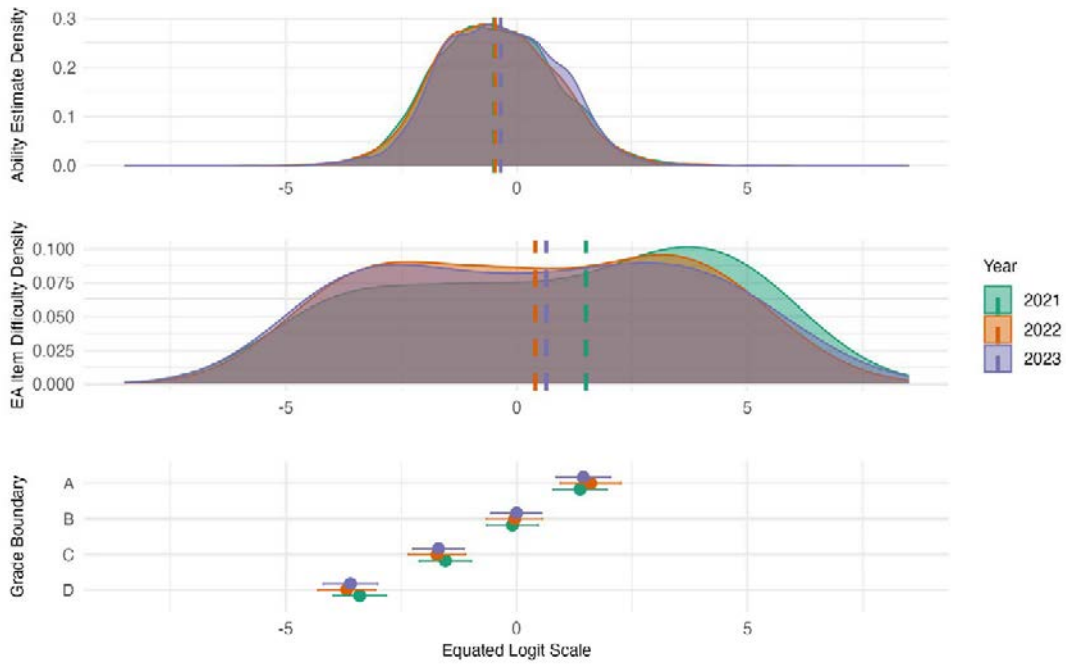


Figure 9.13 Plots for Design showing the distribution and median of the equated ability estimates from the combined IA and EA Rasch analysis (top panel), the distribution and median of the equated EA item difficulty estimates (middle panel) and the equated grade boundary cutoffs with their 95% confidence intervals (bottom panel) across the three years

English

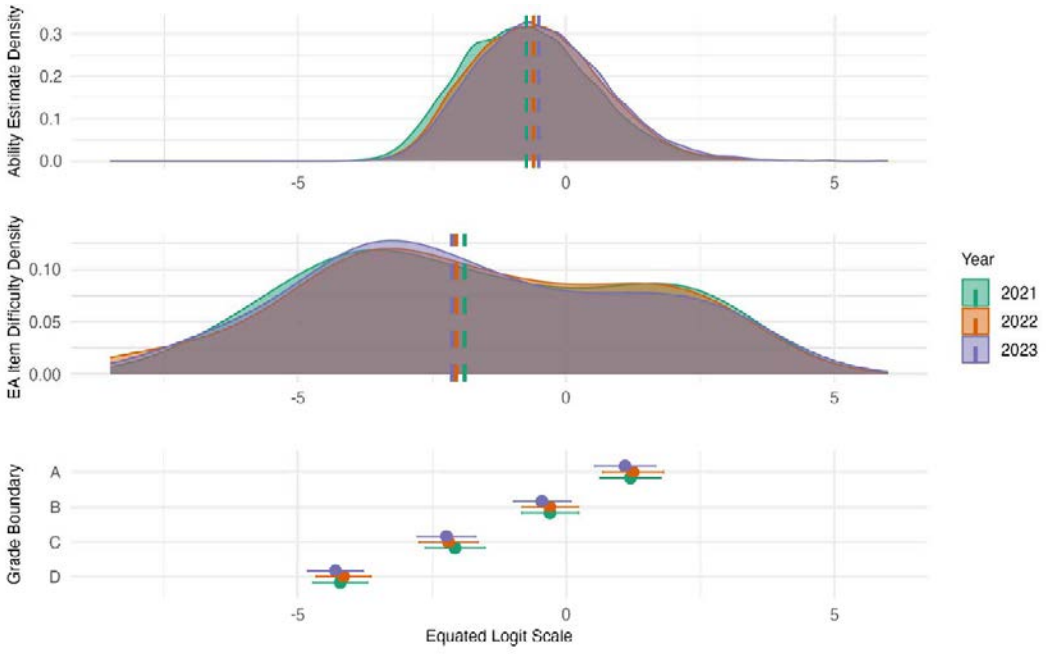


Figure 9.14 Plots for English showing the distribution and median of the equated ability estimates from the combined IA and EA Rasch analysis (top panel), the distribution and median of the equated EA item difficulty estimates (middle panel) and the equated grade boundary cutoffs with their 95% confidence intervals (bottom panel) across the three years

Japanese

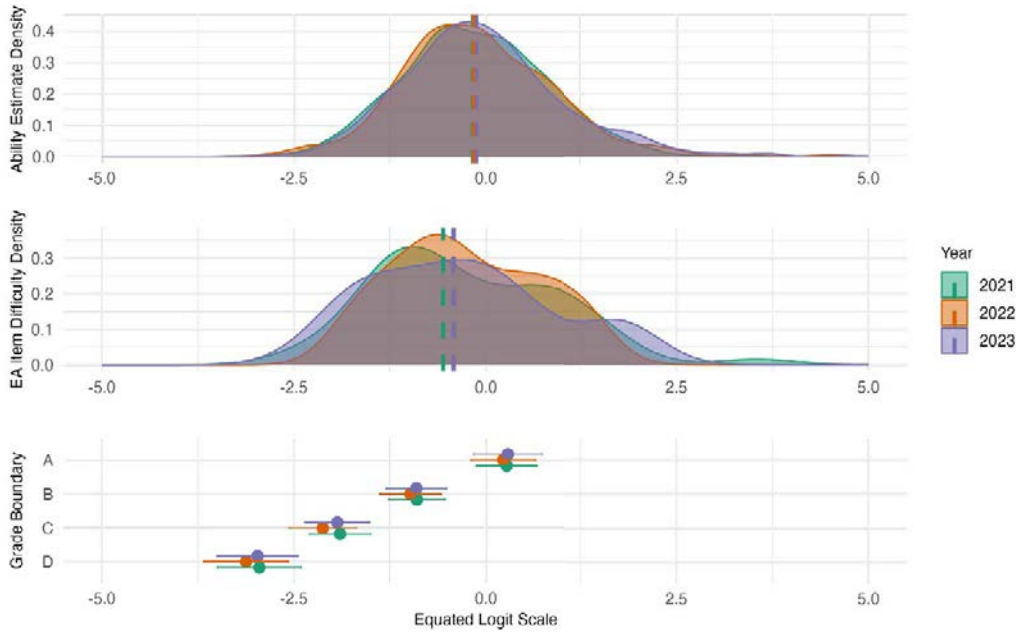


Figure 9.15 Plots for Japanese showing the distribution and median of the equated ability estimates from the combined IA and EA Rasch analysis (top panel), the distribution and median of the equated EA item difficulty estimates (middle panel) and the equated grade boundary cutoffs with their 95% confidence intervals (bottom panel) across the three years

Mathematical Methods

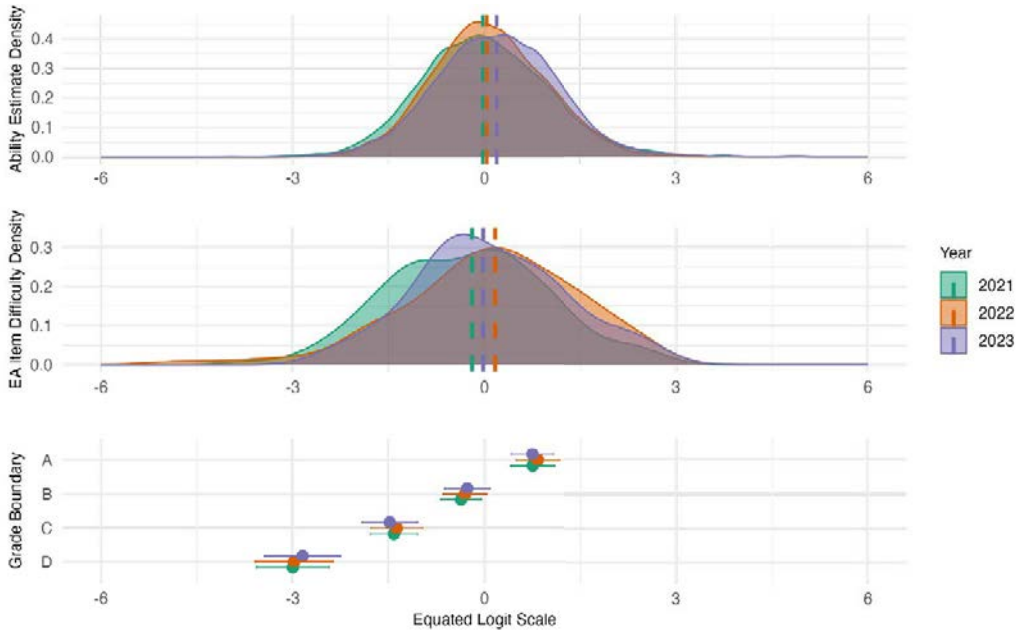


Figure 9.16 Plots for Mathematical Methods showing the distribution and median of the equated ability estimates from the combined IA and EA Rasch analysis (top panel), the distribution and median of the equated EA item difficulty estimates (middle panel) and the equated grade boundary cutoffs with their 95% confidence intervals (bottom panel) across the three years

Modern History

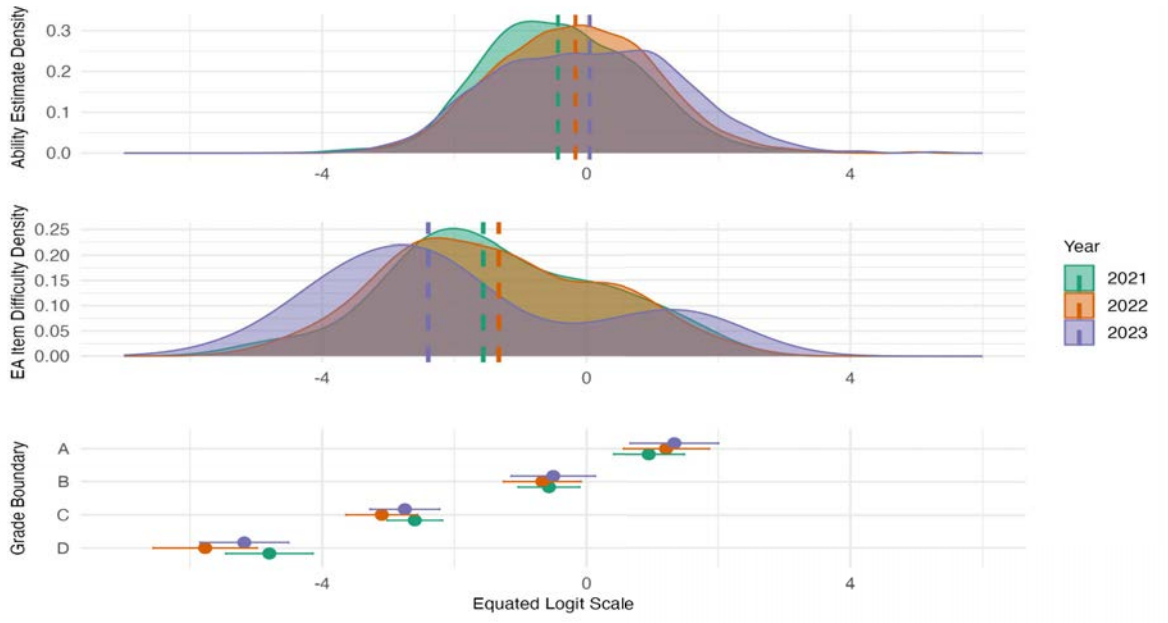


Figure 9.17 Plots for Modern History showing the distribution and median of the equated ability estimates from the combined IA and EA Rasch analysis (top panel), the distribution and median of the equated EA item difficulty estimates (middle panel) and the equated grade boundary cutoffs with their 95% confidence intervals (bottom panel) across the three years

Physical Education

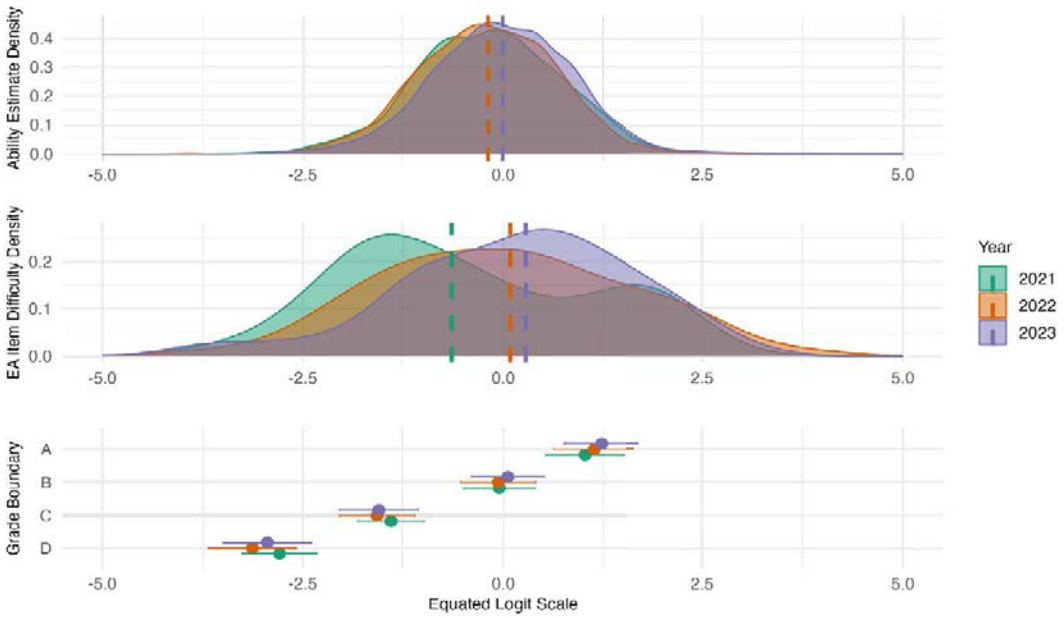


Figure 9.18 Plots for Physical Education showing the distribution and median of the equated ability estimates from the combined IA and EA Rasch analysis (top panel), the distribution and median of the equated EA item difficulty estimates (middle panel) and the equated grade boundary cutoffs with their 95% confidence intervals (bottom panel) across the three years

Visual Art

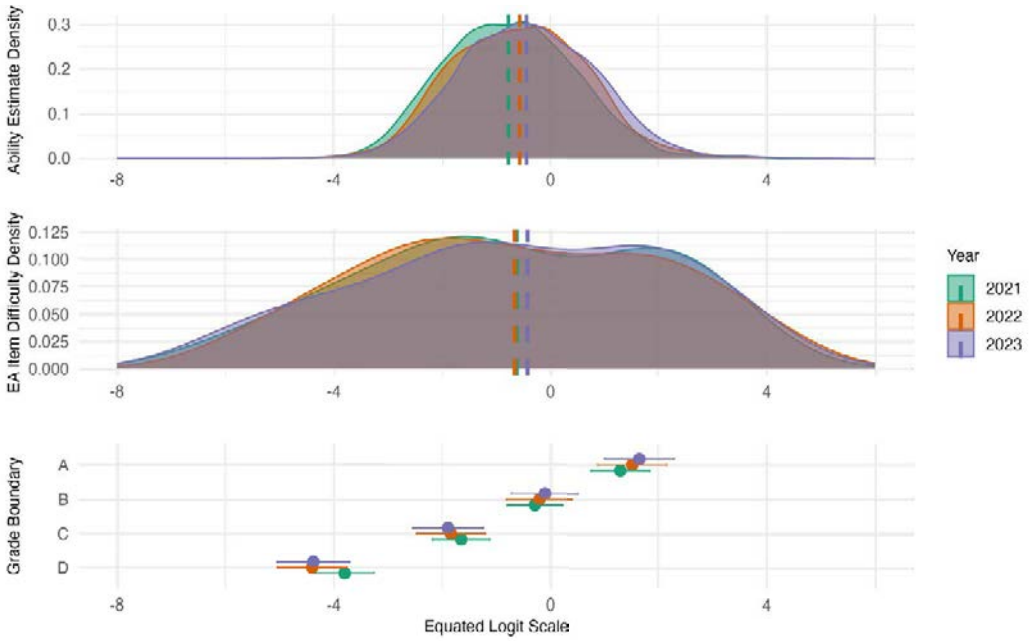


Figure 9.19 Plots for Visual Art showing the distribution and median of the equated ability estimates from the combined IA and EA Rasch analysis (top panel), the distribution and median of the equated EA item difficulty estimates (middle panel) and the equated grade boundary cutoffs with their 95% confidence intervals (bottom panel) across the three year

9.3 Appendix C - Survey Statistics

9.3.1 Summary of surveys

The table below presents information on the surveys administered as part of this evaluation.

1. The first column provides the title of the survey, including an indication of the targeted respondents.
2. The second column indicates how many items in total were on the survey, noting that some of the surveys had branching schemes. This means that, dependent on the specific role or response to items, a respondent could 'branch' into a subset of survey items specifically aligned to their role and/or experience. For example, a Chief Confirmer would be directed to a subset of the survey items about aspects of the system specific to their role, such as quality assurance steps or app functionality. A Confirmer would not see those items as their role does not interact with those features in the same way.
3. The third column indicates the number of people who responded to any item on the survey. Note that responding to all items was not required and respondents were branched in some surveys as noted above, so the number of responses to any individual item on the survey may vary. The number of respondents is referenced in the tables for each individual survey in Appendices 9.3.2–9.3.6 under 'Item response rate'. This value is the number of respondents who answered the specific item divided by the total number of respondents to any item on the survey.
4. The fourth column indicates the sample size. Not all eligible people classified as members of the population were surveyed. This could be for a variety of reasons, including access to current contact information or the cost and complexity of surveying every member of the population (see point 5 for definition of the population). Invitations to respond to the survey were sent to selected members of the overall population so that the respondent pool represented the population on key factors such as role or location. The number of people in this selected group is the sample size. This information is referenced in the tables for each individual survey, in the column labelled 'Sampling rate'. This value is the number of respondents to the specific item divided by the population size.
5. The fifth column indicates the population size. This is the total number of people who were potentially eligible to respond to the survey based on their role or experience with the QCE system. A selected sample of this group was invited to respond to the appropriate survey.

As an example, in the 'Targeted Survey: Confirmation' (row 2) a population of 2,064 individuals have been employed as an assessor by QCAA in some role in Confirmation since the beginning of the new QCE program. These roles include Confirmer, Lead Confirmer and Chief Confirmer, noting that some of the eligible individuals may have been employed in more than one of these roles over time. From these 2,064 people, a sample of 741 people was selected, including 489 Confirmers, 129 Lead Confirmers, and 17 Chief Confirmers. This sample was invited to respond to the survey as part of the evaluation. Of those 741 individuals, 610 participated in the survey by responding to at least one of the 40 items on the survey (details of the item response rates appear in subsequent tables for each survey). Those 610 individuals comprised 480 Confirmers (out of 489 invited, or 98.2%); 114 Lead Confirmers (out of 129 invited, or 88.4%); and 16 Chief Confirmers (out of 179 invited, or 94.1%).

Table 9.5 Summary of surveys

Survey title	Questions	Respondents	Sample size	Population
1. Targeted Survey: Endorsement	21	322	389	843
2. Targeted Survey: Confirmation	40	610 (Total) 480 (Confirmer) 114 (Lead Confirmer) 16 (Chief Confirmer)	741 (Total) 489 (Confirmer) 129 (Lead Confirmer) 17 (Chief Confirmer)	2,064
3. Targeted Survey: QCAA Staff Survey	59	100	140	351
4. Targeted Survey: Invigilators	22	273	339	600
5. Targeted Survey: External Assessment Markers	48	1,028 (Total) 826 (Marker) 175 (Lead Marker) 27 (Chief Marker)	1,167 (Total) 831 (Marker) 178 (Lead Marker) 27 (Chief Marker)	3,121
6. General Survey: QCAA General Survey	39	894	2261	N/A
7. Online School Leader Forum	18	369 (Total) 107 (Principal) 225 (Principal's Delegate) 37 (Other)	370 (Total) 107 (Principal) 225 (Principal's Delegate) 38 (Other)	N/A
	247	**3,596**	**5,407**	**N/A**

9.3.2 Targeted surveys

Table 9.6 Endorsement targeted survey questions and descriptive statistics

No.	Questions	Item respondents	Sample size	Item response rate[1]	Population	Sampling rate
1.01	The pre-event briefing session was useful in preparing me for the Endorsement event.	321	389	82.52%	843	38.08%
1.02	The online modules were useful in developing my knowledge about assessment design and the Endorsement process.	322	389	82.78%	843	38.20%
1.03	The training program was useful in preparing me to be an Endorser/Lead Endorser/Chief Endorser.	320	389	82.26%	843	37.96%
1.04	Overall, I felt prepared to carry out my roles as an Endorser/ Lead Endorser/Chief Endorser.	320	389	82.26%	843	37.96%
1.05	The Endorsement events I participated in were well-organised.	318	389	81.75%	843	37.72%
1.06	The Endorsement process and requirements for my role were clearly explained to me.	318	389	81.75%	843	37.72%
1.07	I had enough time to carry out my assigned duties during the Endorsement event.	318	389	81.75%	843	37.72%
1.08	The assessment evaluation tool was easy to apply when reviewing assessment instruments.	318	389	81.75%	843	37.72%
1.09	I am confident in the accuracy of my Endorsement decisions.	318	389	81.75%	843	37.72%
1.10	The QCAA Endorsement app effectively supported me to complete my tasks as an Endorser/Lead Endorser/Chief Endorser.	318	389	81.75%	843	37.72%

[1] Item response rate: The percentage of respondents who answered a specific question, as opposed to skipping it.

No.	Questions	Item respondents	Sample size	Item response rate	Population	Sampling rate
1.11	I received timely feedback and support from the Lead Endorser/ Chief Endorser/PEO I worked with.	317	389	81.49%	843	37.60%
1.12	Overall, I was able to find help and support when needed during the Endorsement process.	317	389	81.49%	843	37.60%
1.13	I intend to take on the role of an Endorser/Lead Endorser/Chief Endorser in the future.	318	389	81.75%	843	37.72%
1.14	The Endorsement process is efficient.	314	389	80.72%	843	37.25%
1.15	The Endorsement process is rigorous and reliable.	314	389	80.72%	843	37.25%
1.16	The Endorsement process is effective in ensuring assessments align with syllabus specifications.	314	389	80.72%	843	37.25%
1.17	The timing of the Endorsement process suits me.	313	389	80.46%	843	37.13%
1.18	The Endorsement process prevents poor quality assessments from being used in schools.	314	389	80.72%	843	37.25%
1.19	The Endorsement process gives useful direction to schools when assessments are not endorsed.	313	389	80.46%	843	37.13%
1.20	The Endorsement process supports continuous improvement in assessment over the years.	313	389	80.46%	843	37.13%
1.21	The Endorsement process is sustainable from a workforce perspective (i.e., recruiting and retaining teachers as Endorsers).	313	389	80.46%	843	37.13%

Table 9.7 Confirmation targeted survey questions and descriptive statistics

No.	Questions[2]	Item respondents	Sample size	Item response rate	Population	Sampling rate
2.01	The briefing at the beginning of the event was useful in preparing me for the Confirmation event.	610	741	82.32%	2064	29.55%
2.02	The online assessment literacy modules were useful in developing my knowledge about making judgments and the Confirmation process.	610	741	82.32%	2064	29.55%
2.03	The training program was useful in preparing me to be a Confirmer/Lead Confirmer/Chief Confirmer.	608	741	82.05%	2064	29.46%
2.04	Overall, I felt prepared to carry out my roles as a Confirmer/Lead Confirmer/Chief Confirmer.	610	741	82.32%	2064	29.55%
2.05	The Confirmation events I participated in were well-organised.	599	741	80.84%	2064	29.02%
2.06	The Confirmation process and requirements for my role were clearly explained to me.	599	741	80.84%	2064	29.02%
2.07	I had enough time to carry out my assigned duties during the Confirmation event.	600	741	80.97%	2064	29.07%
2.08	The process of using an annotated ISMG to review samples is effective.	599	741	80.84%	2064	29.02%
2.09	I am confident in the accuracy of my Confirmation decisions.	597	741	80.57%	2064	28.92%
2.10	The QCAA Confirmation app in the QCAA Portal effectively enabled me to complete my tasks as a Confirmer/Lead Confirmer/Chief Confirmer.	598	741	80.70%	2064	28.97%
2.11	I received timely support and responses to my queries from the Lead Confirmer/Chief Confirmer/PEO I worked with.	599	741	80.84%	2064	29.02%
2.12	Overall, I was able to find help and support when needed during the Confirmation process.	599	741	80.84%	2064	29.02%
2.13	I intend to take on the role of a Confirmer/Lead Confirmer/Chief Confirmer in the future.	599	741	80.84%	2064	29.02%
2.14	The Confirmation events I participated in were well-organised.	114	129	88.37%	2064	5.52%[1]
2.15	The Confirmation process and requirements for my role were clearly explained to me.	113	129	87.60%	2064	5.47%
2.16	I had enough time to carry out my assigned duties during the Confirmation event.	113	129	87.60%	2064	5.47%
2.17	The process of using an annotated ISMG to review samples is effective.	112	129	86.82%	2064	5.43%
2.18	I am confident in the accuracy of my Confirmation decisions.	110	129	85.27%	2064	5.33%

[2] Some survey items were targeted to specific groups (e.g., Lead Confirmers, Chief Confirmers).

No.	Questions	Item respondents	Sample size	Item response rate	Population	Sampling rate
2.19	The QCAA Confirmation app in the QCAA Portal effectively enabled me to complete my tasks as a Confirmer/Lead Confirmer/Chief Confirmer.	112	129	86.82%	2064	5.43%
2.20	I received timely support and responses to my queries from the Lead Confirmer/Chief Confirmer/ PEO I worked with.	112	129	86.82%	2064	5.43%
2.21	Overall, I was able to find help and support when needed during the Confirmation process.	111	129	86.05%	2064	5.38%
2.22	I intend to take on the role of a Confirmer/Lead Confirmer/Chief Confirmer in the future.	112	129	86.82%	2064	5.43%
2.23	The number of Confirmers in my Confirmer group was manageable.	112	129	86.82%	2064	5.43%
2.24	I had sufficient time to complete a lead review of the submissions allocated to my group.	112	129	86.82%	2064	5.43%
2.25	The Confirmation events I participated in were well-organised.	16	17	94.12%	2064	0.78%
2.26	The Confirmation process and requirements for my role were clearly explained to me.	16	17	94.12%	2064	0.78%
2.27	I had enough time to carry out my assigned duties during the Confirmation event.	16	17	94.12%	2064	0.78%
2.28	The process of using an annotated ISMG to review samples is effective.	16	17	94.12%	2064	0.78%
2.29	I am confident in the accuracy of my Confirmation decisions.	16	17	94.12%	2064	0.78%
2.30	The QCAA Confirmation app in the QCAA Portal effectively enabled me to complete my tasks as a Confirmer/Lead Confirmer/Chief Confirmer.	16	17	94.12%	2064	0.78%
2.31	I received timely support and responses to my queries from the Lead Confirmer/Chief Confirmer/ PEO I worked with.	16	17	94.12%	2064	0.78%
2.32	Overall, I was able to find help and support when needed during the Confirmation process.	15	17	88.24%	2064	0.73%
2.33	I intend to take on the role of a Confirmer/Lead Confirmer/Chief Confirmer in the future.	16	17	94.12%	2064	0.78%
2.34	The process to resolve flagged samples was straightforward.	15	17	88.24%	2064	0.73%
2.35	The Confirmation process is efficient.	591	741	79.76%	2064	28.63%
2.36	The Confirmation process is rigorous and reliable.	591	741	79.76%	2064	28.63%
2.37	The Confirmation process is effective in checking that teachers have used the ISMG accurately and consistently when marking student assessments.	591	741	79.76%	2064	28.63%
2.38	The timing of the Confirmation meetings suits me.	591	741	79.76%	2064	28.63%
2.39	The Confirmation process is an essential quality assurance process in the QCE system.	591	741	79.76%	2064	28.63%
2.40	The Confirmation process is sustainable from a workforce perspective (i.e., recruiting and retaining teachers as Confirmers).	591	741	79.76%	2064	28.63%

Table 9.8 QCAA Staff targeted survey questions and descriptive statistics

No.	Questions	Item respondents	Sample size	Item response rate	Population	Sampling rate
3.01	At the QCAA, we: enjoy discussing strategies for monitoring and evaluating.	100	140	71.43%	351	28.49%
3.02	At the QCAA, we are eager to engage in monitoring and evaluation.	100	140	71.43%	351	28.49%
3.03	At the QCAA, we: believe evaluation is a valuable endeavour.	100	140	71.43%	351	28.49%
3.04	At the QCAA, we: question assumptions or claims made by each other.	100	140	71.43%	351	28.49%
3.05	At the QCAA, we: offer evidence for claims that we make.	100	140	71.43%	351	28.49%
3.06	At the QCAA, we: suggest alternative explanations.	100	140	71.43%	351	28.49%
3.07	At the QCAA, we: take time to reflect on the way we work.	100	140	71.43%	351	28.49%
3.08	QCAA has a strong organisational culture that supports staff members' learning and development.	100	140	71.43%	351	28.49%
3.09	Evidence-informed decision making is encouraged at the QCAA.	100	140	71.43%	351	28.49%
3.10	I can access monitoring information relevant to the decisions I need to make in my role.	100	140	71.43%	351	28.49%
3.11	The data I access for monitoring within my role is in a format that is easy to use.	100	140	71.43%	351	28.49%
3.12	I am satisfied with the change management process at the QCAA.	99	140	70.71%	351	28.21%
3.13	I am satisfied with the communication process at the QCAA.	100	140	71.43%	351	28.49%
3.14	I am satisfied with the way leaders at the QCAA have managed and communicated changes occurring at the QCAA.	100	140	71.43%	351	28.49%
3.15	I am confident that the QCAA's ICT systems will be adequate to support the future needs of my role.	100	140	71.43%	351	28.49%

No.	Questions	Item respondents	Sample size	Item response rate	Population	Sampling rate
3.16	I am confident that the new QCE system is fair and equitable for students across Queensland.	90	140	64.29%	351	25.64%
3.17	I am confident that the new QCE system is a high-quality certification system.	90	140	64.29%	351	25.64%
3.18	I am confident I can implement the requirements of the new QCE system associated with my role.	89	140	63.57%	351	25.36%
3.19	I am confident the QCAA can fulfil its responsibilities in relation to the implementation of the new QCE system.	90	140	64.29%	351	25.64%
3.20	The QCAA can adapt the QCE design to meet changing circumstances.	90	140	64.29%	351	25.64%
3.21	I understand how my role contributes to the effectiveness of the new QCE system.	90	140	64.29%	351	25.64%
3.22	I have received adequate training to enact my role in implementing the new QCE system.	90	140	64.29%	351	25.64%
3.23	I feel well prepared to implement my role in the new QCE system.	90	140	64.29%	351	25.64%
3.24	I feel positive about my role in implementing the new QCE system.	90	140	64.29%	351	25.64%
3.25	The QCAA's ICT systems that I use are adequate to support the ongoing needs of my role.	90	140	64.29%	351	25.64%
3.26	The QCAA's ICT system has helped me greatly in my role related to the implementation of the new QCE system.	90	140	64.29%	351	25.64%
3.27	The monitoring processes at the QCAA capture information that is valuable to inform decision-making about the new QCE system.	90	140	64.29%	351	25.64%
3.28	The QCAA gathers evidence that can be used to improve the new QCE system.	90	140	64.29%	351	25.64%
3.29	The QCAA provides professional learning opportunities to assist schools to access the data they need.	90	140	64.29%	351	25.64%

No.	Questions	Item respondents	Sample size	Item response rate	Population	Sampling rate
3.30	The QCAA assessors and casual assessors within the new QCE system understand their roles.	50	62	80.65%	351	14.25%
3.31	The QCAA assessors and casual assessors within the new QCE system implement their roles consistently.	50	62	80.65%	351	14.25%
3.32	The QCAA assessors and casual assessors are positive about implementing their roles.	50	62	80.65%	351	14.25%
3.33	I am confident that the new QCE system provides high-quality assessments that prepare students for post-school pathways.	50	62	80.65%	351	14.25%
3.34	The QCAA's ICT systems support the timely completion of assessment activities.	50	62	80.65%	351	14.25%
3.35	I am confident that the QCAA processes support the sustainability of the assessor workforce.	50	62	80.65%	351	14.25%
3.36	People teaching the new QCE curriculum understand its requirements.	35	42	83.33%	351	9.97%
3.37	Teachers implement the new QCE curriculum in accordance with the curriculum documents.	35	42	83.33%	351	9.97%
3.38	Teachers are confident delivering the new QCE curriculum.	35	42	83.33%	351	9.97%
3.39	The new QCE curriculum prepares students for their post-school life.	35	42	83.33%	351	9.97%
3.40	The new QCE curriculum develops students' 21st century skills.	35	42	83.33%	351	9.97%
3.41	Teachers and schools I have contact with are satisfied with the curriculum resources and supports the QCAA offers.	35	42	83.33%	351	9.97%
3.42	Schools (including QCAA assessors) know how to use the QCAA's ICT Portal apps related to their roles.	18	20	90.00%	351	5.13%
3.43	Schools (including QCAA assessors) use the QCAA's ICT Portal apps effectively.	18	20	90.00%	351	5.13%
3.44	Schools (including QCAA assessors) are confident using the QCAA's ICT Portal apps.	18	20	90.00%	351	5.13%

No.	Questions	Item respondents	Sample size	Item response rate	Population	Sampling rate
3.45	Continuous improvements are being made to the QCAA's ICT system based on feedback from schools and/or QCAA assessors.	18	20	90.00%	351	5.13%
3.46	The ICT Portal apps make it easy for schools and/or QCAA assessors to access the data they need to monitor implementation and progress.	18	20	90.00%	351	5.13%
3.47	Schools implement AARA processes in accordance with the documented processes.	8	14	57.14%	351	2.28%
3.48	Schools are confident in their ability to implement AARA processes.	8	14	57.14%	351	2.28%
3.49	The AARA processes and requirements are adaptable to the different school contexts.	8	14	57.14%	351	2.28%
3.50	The AARA processes and requirements are adaptable to the different student needs.	8	14	57.14%	351	2.28%
3.51	The teachers and schools I have contact with find the QCAA's AARA resources useful.	8	14	57.14%	351	2.28%
3.52	The QCAA provides high quality professional development/training to schools/QCAA assessors.	40	44	90.91%	351	11.40%
3.53	The QCAA professional development and resources keep pace with changes within the new QCE system.	40	44	90.91%	351	11.40%
3.54	The professional development and training the QCAA provides prepares schools/QCAA assessors to enact their roles.	40	44	90.91%	351	11.40%
3.55	The QCAA provides the types of professional development/training that schools/QCAA assessors need.	40	44	90.91%	351	11.40%
3.56	The QCAA provides professional development/training in the modes that schools/QCAA assessors prefer.	40	44	90.91%	351	11.40%
3.57	The QCAA professional development and resources provided to schools/QCAA assessors support them to access data via the ICT Portal apps.	40	44	90.91%	351	11.40%
3.58	Teachers and schools I have contact with find the QCAA's professional development activities useful.	40	44	90.91%	351	11.40%
3.59	QCAA's HR and finance processes are effective at supporting the assessor workforce	15	17	88.24%	351	4.27%

Table 9.9 Invigilators targeted survey questions and descriptive statistics

No.	Questions	Item respondents	Sample size	Item response rate	Population	Sampling rate
4.01	The application process to become an invigilator was easy and straightforward.	272	339	80.24%	600	45.33%
4.02	The *Handbook for Invigilators* was useful in preparing me for my role as an invigilator.	273	339	80.53%	600	45.50%
4.03	The online training program was useful in preparing me to be an invigilator.	273	339	80.53%	600	45.50%
4.04	Overall, I felt prepared to carry out my roles as an invigilator.	273	339	80.53%	600	45.50%
4.05	I was able to contact each assessment venue/school with minimal issues before the assessment period.	263	339	77.58%	600	43.83%
4.06	I was able to communicate with the EA supervisors with minimal issues on the day.	269	339	79.35%	600	44.83%
4.07	I was informed by the EA supervisor/ school staff when an incident occurred during the assessment session.	163	339	48.08%	600	27.17%
4.08	I was aware of actions I needed to take if issues or violations occurred during an assessment event.	260	339	76.70%	600	43.33%
4.09	I received timely support and responses to my queries from QCAA during the assessment sessions.	121	339	35.69%	600	20.17%
4.10	Overall, I was happy with the way QCAA dealt with issues related to invigilation that I raised.	119	339	35.10%	600	19.83%
4.11	I was able to conduct my role as an invigilator independently.	271	339	79.94%	600	45.17%
4.12	The number of assessment sessions I had to invigilate each day was manageable.	269	339	79.35%	600	44.83%

No.	Questions	Item respondents	Sample size	Item response rate	Population	Sampling rate
4.13	The students at the school I invigilated appeared to be aware of the External Assessment rules and directions.	269	339	79.35%	600	44.83%
4.14	The students I invigilated followed the rules of External Assessment.	270	339	79.65%	600	45.00%
4.15	The invigilator checklist was useful in helping me record information about each session.	269	339	79.35%	600	44.83%
4.16	I had enough time to complete and submit my online invigilator report at the end of each assessment day.	270	339	79.65%	600	45.00%
4.17	The QCAA Portal effectively enabled me to complete my tasks as invigilator.	270	339	79.65%	600	45.00%
4.18	Overall, I was able to find help and support when needed during the External Assessment period.	191	339	56.34%	600	31.83%
4.19	I intend to apply to be an invigilator in the future.	266	339	78.47%	600	44.33%
4.20	I understood my main responsibilities as an invigilator.	270	339	79.65%	600	45.00%
4.21	The invigilation process contributes to public confidence in the QCE system.	270	339	79.65%	600	45.00%
4.22	The advice and support provided by QCAA have been consistent.	268	339	79.06%	600	44.67%

Table 9.10 External Assessment markers targeted survey questions and descriptive statistics

No.	Questions	Item respondents	Sample size	Item response rate	Population	Sampling rate
5.01	The application process to be a Marker was easy and straightforward.	1028	1167	88.09%	3121	32.94%
5.02	The online assessment literacy modules were useful in developing my knowledge about making judgments as a Marker.	1027	1167	88.00%	3121	32.91%
5.03	The training program I completed was useful in preparing me to be a Marker/ Lead Marker/Chief Marker.	1027	1167	88.00%	3121	32.91%
5.04	Overall, I felt prepared to carry out my role as a Marker/Lead Marker/Chief Marker.	1028	1167	88.09%	3121	32.94%
5.05	The External Assessment marking operation was well-organised.	794	831	95.55%	3121	25.44%
5.06	The marking process and requirements for my role were clearly explained to me.	795	831	95.67%	3121	25.47%
5.07	The External Assessment marking guide (EAMG) was easy to understand and use.	795	831	95.67%	3121	25.47%
5.08	The recalibration processes embedded into the marking operation have helped me to become a more accurate Marker.	709	831	85.32%	3121	22.72%
5.09	I am confident in the accuracy of my marking decisions.	795	831	95.67%	3121	25.47%
5.10	The MarkManager online platform effectively enabled me to complete my tasks as a Marker.	795	831	95.67%	3121	25.47%
5.11	The expected rate of marking is achievable.	791	831	95.19%	3121	25.34%
5.12	I received timely support and responses to my queries from the Lead Marker/Chief Marker/PEO I worked with.	749	831	90.13%	3121	24.00%

No.	Questions	Item respondents	Sample size	Item response rate	Population	Sampling rate
5.13	I was given consistent advice on how to mark.	771	831	92.78%	3121	24.70%
5.14	Overall, I was able to find help and support when needed during the marking period.	760	831	91.46%	3121	24.35%
5.15	I intend to take on the role of a Marker/Lead Marker/Chief Marker in the future.	783	831	94.22%	3121	25.09%
5.16	I was able to access useful data to help me track my performance as a Marker.	771	831	92.78%	3121	24.70%
5.17	The External Assessment marking operation was well-organised.	175	178	98.31%	3121	5.61%
5.18	The marking process and requirements for my role were clearly explained to me.	175	178	98.31%	3121	5.61%
5.19	I had enough time to prepare for Marker training.	175	178	98.31%	3121	5.61%
5.20	The script selection process was efficient.	175	178	98.31%	3121	5.61%
5.21	I had enough time to mark all scripts required.	173	178	97.19%	3121	5.54%
5.22	The number of Markers allocated to me was manageable.	173	178	97.19%	3121	5.54%
5.23	The process to re-calibrate Markers was clear and straightforward.	172	178	96.63%	3121	5.51%
5.24	The process to re-calibrate Markers was effective.	164	178	92.13%	3121	5.25%
5.25	I am confident in the accuracy of my marking decisions.	174	178	97.75%	3121	5.58%
5.26	The MarkManager online platform effectively enabled me to complete my tasks as a Lead Marker.	173	178	97.19%	3121	5.54%
5.27	I received timely support and responses to my queries from the Chief Marker I worked with.	172	178	96.63%	3121	5.51%

No.	Questions	Item respondents	Sample size	Item response rate	Population	Sampling rate
5.28	Overall, I was able to find help and support when needed during the marking period.	172	178	96.63%	3121	5.51%
5.29	I intend to take on the role of a Marker/ Lead Marker/Chief Marker in the future.	173	178	97.19%	3121	5.54%
5.30	The External Assessment marking operation was well-organised.	27	27	100.00%	3121	0.87%
5.31	The marking process and requirements for my role were clearly explained to me.	27	27	100.00%	3121	0.87%
5.32	The script selection process was efficient.	27	27	100.00%	3121	0.87%
5.33	The number of Markers/Lead Markers allocated to me was manageable.	27	27	100.00%	3121	0.87%
5.34	The process to re-calibrate Markers was clear and straightforward.	27	27	100.00%	3121	0.87%
5.35	The process to re-calibrate Markers was effective.	27	27	100.00%	3121	0.87%
5.36	I am confident in the accuracy of my marking decisions.	27	27	100.00%	3121	0.87%
5.37	The MarkManager online platform effectively enabled me to complete my tasks as a Chief Marker.	27	27	100.00%	3121	0.87%
5.38	I was given clear guidelines on how to manage the reassessment process.	22	27	81.48%	3121	0.70%
5.39	I received timely support and responses to my queries from the PEO I worked with.	27	27	100.00%	3121	0.87%
5.40	Overall, I was able to find help and support when needed during the marking period.	27	27	100.00%	3121	0.87%
5.41	I intend to take on the role of a Marker/ Lead Marker/Chief Marker in the future.	27	27	100.00%	3121	0.87%
5.42	The External Assessment marking process is efficient.	986	1167	84.49%	3121	31.59%

No.	Questions	Item respondents	Sample size	Item response rate	Population	Sampling rate
5.43	The marking process is rigorous and reliable.	985	1167	84.40%	3121	31.56%
5.44	I am confident about the security measures put in place during the marking operations.	978	1167	83.80%	3121	31.34%
5.45	The External Assessment process is transparent and fair.	978	1167	83.80%	3121	31.34%
5.46	The hours allowed for marking operations suit me.	989	1167	84.75%	3121	31.69%
5.47	The timing of the reassessment period suits me.	873	1167	74.81%	3121	27.97%
5.48	The External Assessment marking process is sustainable from a workforce perspective (i.e., recruiting eligible Markers annually).	944	1167	80.89%	3121	30.25%

9.3.3 General survey

Table 9.11 QCAA general survey questions and descriptive statistics

No.	Questions	Item respondents	Sample size	Item response rate	Population	Sampling rate
6.01	I have a good understanding of the new QCE system.	894	2261	39.54%	*	*
6.02	The new QCE system fully addresses the needs of: teachers	888	2261	39.27%	*	*
6.03	The new QCE system fully addresses the needs of: students	883	2261	39.05%	*	*
6.04	The new QCE system fully addresses the needs of: higher education providers	875	2261	38.70%	*	*
6.05	The new QCE system fully addresses the needs of: employers and industry	876	2261	38.74%	*	*
6.06	The new QCE system fully addresses the needs of: community members	872	2261	38.57%	*	*
6.07	The new QCE system is fair and equitable for my students.	891	2261	39.41%	*	*
6.08	I have a good understanding of the QCAA's responsibilities implementing the new QCE system.	890	2261	39.36%	*	*
6.09	QCAA fulfils its responsibilities implementing the new QCE.	888	2261	39.27%	*	*
6.10	The resources and support offered by QCAA enable me to effectively: Write and plan curriculum	845	2261	37.37%	*	*
6.11	The resources and support offered by QCAA enable me to effectively: Write Internal Assessments	833	2261	36.84%	*	*
6.12	The resources and support offered by QCAA enable me to effectively: Have Internal Assessments endorsed	823	2261	36.40%	*	*
6.13	The resources and support offered by QCAA enable me to effectively: Have Internal Assessment results confirmed	824	2261	36.44%	*	*
6.14	The resources and support offered by QCAA enable me to effectively: Prepare students for External Assessments	841	2261	37.20%	*	*

**Data not available*

No.	Questions	Item respondents	Sample size	Item response rate	Population	Sampling rate
6.16	I have a good understanding of what the various roles do in the new QCE system (e.g., principal's delegates, syllabus and exam writers, Endorsers, Confirmers and Markers).	871	2261	38.52%	*	*
6.17	The implementation of the new QCE system is progressing well at my school.	870	2261	38.48%	*	*
6.18	How consistent do you think the QCE system implementation is across different school contexts (e.g. school location, school sector)?	804	2261	35.56%	*	*
6.19	How satisfied were you with the professional development you participated in?	788	2261	34.85%	*	*
6.20	At the beginning of your involvement with the new QCE system, how prepared did you feel to undertake your role?	837	2261	37.02%	*	*
6.21	In 2024, how prepared do you feel to undertake your role in the new QCE system?	837	2261	37.02%	*	*
6.22	How satisfied are you with the QCAA's ICT support (e.g., client services, quick step guides and other resources) in the new QCE system?	669	2261	29.59%	*	*
6.23	The QCAA ICT systems/apps help me in my role in the new QCE.	828	2261	36.62%	*	*
6.24	How suitable is the design of the ICT system for managing access arrangements and reasonable adjustments (AARA)?	172	428	40.19%	*	*
6.25	How easy is it for staff at your school to access the data they need via the QCAA Portal?	185	428	43.22%	*	*
6.26	It is easy for me to understand the data and reports that are provided by the QCAA.	837	2261	37.02%	*	*
6.27	It is easy for me to use the data and reports provided by the QCAA to inform my teaching practice (including planning and decision making).	834	2261	36.89%	*	*
6.28	Planning has taken place at my school to sustain the implementation of the new QCE system.	836	2261	36.97%	*	*

**Data not available*

No.	Questions	Item respondents	Sample size	Item response rate	Population	Sampling rate
6.29	I think we have adequate staff and/or materials to implement the QCE system.	184	428	42.99%	*	*
6.30	The leaders at my school are supportive of my efforts to implement the QCE system.	183	428	42.76%	*	*
6.31	I think the QCE system addresses a significant need.	183	428	42.76%	*	*
6.32	Providing quality senior secondary assessment is a priority for me.	183	428	42.76%	*	*
6.33	Providing quality senior secondary assessment is a priority for my school.	184	428	42.99%	*	*
6.34	I think that the QCE system leads to positive outcomes for students.	183	428	42.76%	*	*
6.35	The QCE system is better than the previous system.	177	428	41.36%	*	*
6.36	I feel prepared to continue to implement what I learned at professional development sessions.	183	428	42.76%	*	*
6.37	I have the knowledge and skills necessary to continue to implement the QCE system.	184	428	42.99%	*	*
6.38	I don't want to change my approach to assessment from what I was doing prior to the new QCE system.	170	428	39.72%	*	*
6.39	I am quick to embrace new assessment methods.	179	428	41.82%	*	*

**Data not available*

9.3.4 Online School Leader Forum survey

Table 9.12 Online school leader forum survey questions and descriptive statistics

No.	Questions	Item respondents	Sample size	Item response rate	Population	Sampling rate
7.01	Does the QCE curriculum adequately develop students' 21st century skills?	346	370	93.51%	*	*
7.02	Has the Endorsement/Confirmation process led to better internal assessments?	346	370	93.51%	*	*
7.03	Does the QCE suit your students and their contexts?	327	370	88.38%	*	*
7.04	Is the balance of internal and external summative assessment within the QCE right?	345	370	93.24%	*	*
7.05	Should the Endorsement/Confirmation process to be extended to applied subjects?	344	370	92.97%	*	*
7.06	Do four summative assessments allow for a fair indication of what students know and can do?	341	370	92.16%	*	*
7.07	Are the QCAA's curriculum resources and professional development useful?	341	370	92.16%	*	*
7.08	Has the assessment literacy of your teachers improved as a result of the new QCE?	340	370	91.89%	*	*
7.09	Are you finding the new QCE system manageable to implement in your school?	369	370	99.73%	*	*
7.10	Could the QCAA make the QCE easier to implement without compromising quality?	362	370	97.84%	*	*
7.11	Is the professional development provided by the QCAA sufficient to implement the QCE?	367	370	99.19%	*	*
7.12	Is the QCAA Portal user friendly?	366	370	98.92%	*	*
7.13	Overall, is the new QCE is better than the previous system?	295	370	79.73%	*	*
7.14	Are your students better off under the new QCE?	344	370	92.97%	*	*
7.15	Does the QCE suit your students and their contexts?	344	370	92.97%	*	*
7.16	Does the QCAA Portal provide you and your teachers the information you need?	344	370	92.97%	*	*
7.17	Can your school do everything required to implement the QCE in 2024?	342	370	92.43%	*	*
7.18	Is the QCAA responsive to feedback from schools?	345	370	93.24%	*	*

**Data not available*

9.4 Appendix D - Interview protocol and observation note-taking template

9.4.1 Sample of interview protocol [1]

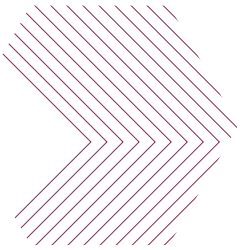
Endorsement Focus Group Protocol

Interviewer introduction notes

- Remind all participants of the voluntary nature of their participation. Describe the 48-hour cooling-off period during which they can withdraw their data before processing begins.
- Ensure that confidentiality protocols as well as limits to confidentiality are explained, provide opportunity for the participants to ask questions.
- Confirm informed consent has been obtained and commence audio recording. PLS and Consent survey link https://q.surveys.unimelb.edu.au/jfe/form/SV_3eEMUlNTPEeBLrE

Interviewer: *Thank you for agreeing to part of this focus group by principals in other parts of our data collection for the QCE evaluation. Before we dive into these areas, could you please tell me a little about your roles in relation to Endorsement.*

[1] A total of 13 focus group interviews were conducted by academics at the University of Melbourne (UoM). The focus groups followed a standardised protocol to ensure consistency in data collection. Although the exact wording or emphasis of some questions was adapted to suit the participants' specific roles or contexts, the overall structure, themes and objectives of the interviews remained consistent across all sessions.



TOR coding	Question	Note
Preparedness for introduction of Endorsement 2.1.1 2.2.0 2.1.2	Please think back to the preparation for the introduction of the QCE. • How well prepared was the QCAA for the introduction of Endorsement? What did QCAA learn during the initial Endorsement activities? • How easily did schools handle the introduction of Endorsement? • How has the preparedness of schools and teachers changed from the initial roll out until today?	
Suitability of current Endorsement processes 4.1.2 5.2.0 5.3.0	Let's look at the current situation with Endorsement. What is working well and what needs to be improved? • Effectiveness of process • Endorsement app • Consistency of training across subjects • Consistency of Endorsement decisions within and across subjects • Support for Endorsers • Time allocation for Endorsers • Collaboration across QCAA teams o Responsibilities, lines of communication and decision making o Documentation • Effectiveness of feedback to schools	
Sustainability of the Endorser workforce 4.2.1 6.3.1 6.2.0	How sustainable is the current external Endorsement workforce? • What workforce planning is in place to support the sustainability of the workforce? • Data use for tracking? • What could be done to improve the sustainability? How sustainable is the Endorsement process for internal staff?	
Organisational change 6.1.2 5.2.0	What processes are in place to communicate changes in interpretation of the assessment requirements to schools and teachers? How effective are these processes? How well do communication channels within QCAA function to keep everyone up to date with changes to Endorsement? How much change to you foresee in the Endorsement process in the coming years? What changes do you anticipate? How feasible are these?	
Overall	Is there anything else you would like to share that you think is important for the evaluation of the QCE?	

9.4.2 Sample of observation note-taking template[2]

Name of event:	Meeting name here
Date:	Choose a date
Time:	[00:00 AM/PM]
Location:	[Location Name]
Mode of observation:	Online or In-person or Hybrid (Please specify)
Details of attendees:	[Names]
Number of attendees:	
Details of trainers/facilitators:	
Number of trainers/facilitators:	

General notes about the event

- Objectives of the event:
- Key discussions:
- Issues/challenges encountered:
- Issues mitigated or resolved:

[2] At sessions observed by academics at the University of Melbourne (UoM) a consistent note-taking template was used. Although the interpretation or appropriateness of some items was adapted to suit the specific contexts, the overall structure of the notes taken was consistent across all sessions.

Fidelity of implementation (Consistency)

1.1 **How consistent was the implementation of processes/instructions observed during the event?**

[Insert key observation #1]

[Details]

[Insert key observation #2]

[Details]

1.2 **Information regarding patterns across school contexts discussed at the event**

[Insert topic discussed]

[Details]

[Insert topic discussed]

[Details]

1.3 **Other observations about consistency:**

[Details]

Implementation progress (Attitude, confidence and readiness of stakeholders)

1.4 **How confident and ready were the trainers or facilitators of the events?**

1.5 **How confident and ready were the participants?**

1.6 **To what extent were participants/trainers satisfied with the ICT system and ICT support received in relation to the process?**

1.7 **Other details about attitude, confidence and readiness of participants/trainers**

Effectiveness (Curriculum support aligns with QCE objectives)

1.8 **To what extent were participants satisfied with the curriculum resources and support received in relation to the process?**

Monitoring (Access to data)

1.9 **How easy was it for participants to access data needed to complete the process?**

9.5 Appendix E - Survey figures

Chapter 3 Findings: Foundation concepts and system developmet

Figure 9.20 Endorsers' perspectives on sustainability and improvement

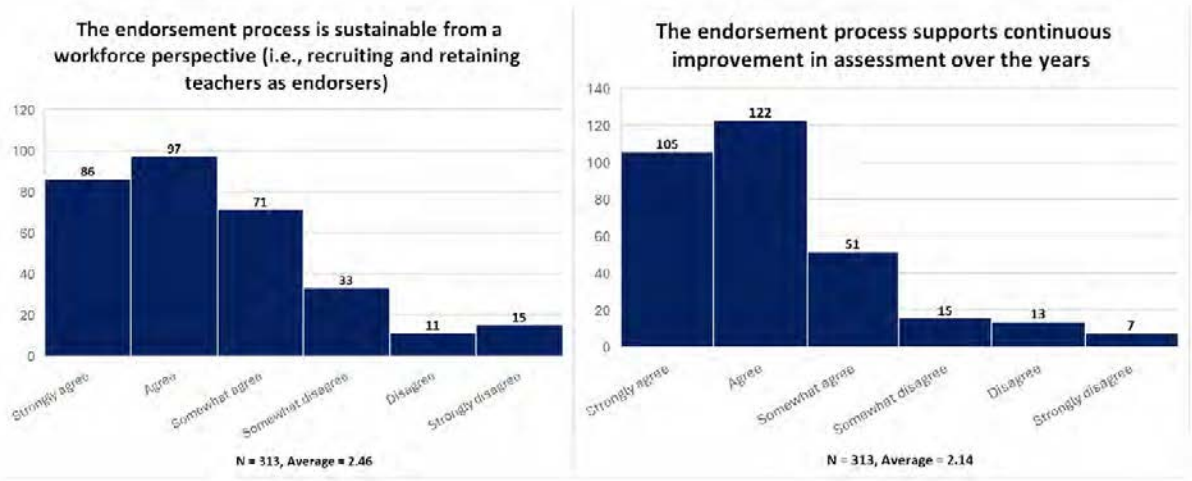


Figure 9.21 General survey reports of the new QCE system

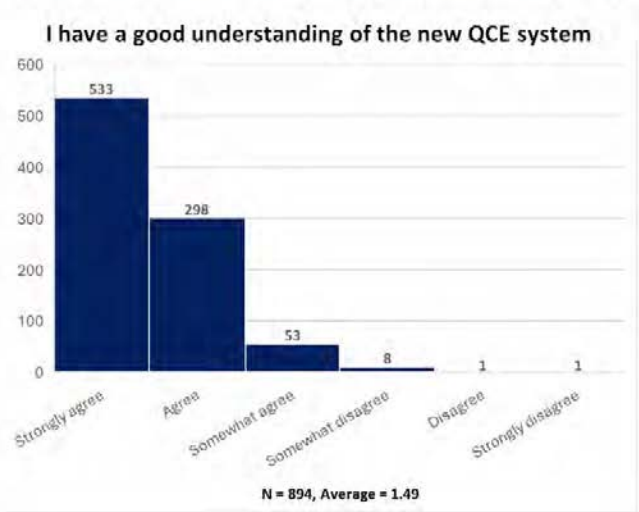


Figure 9.22 Endorsers' perspectives of sustainability and improvement

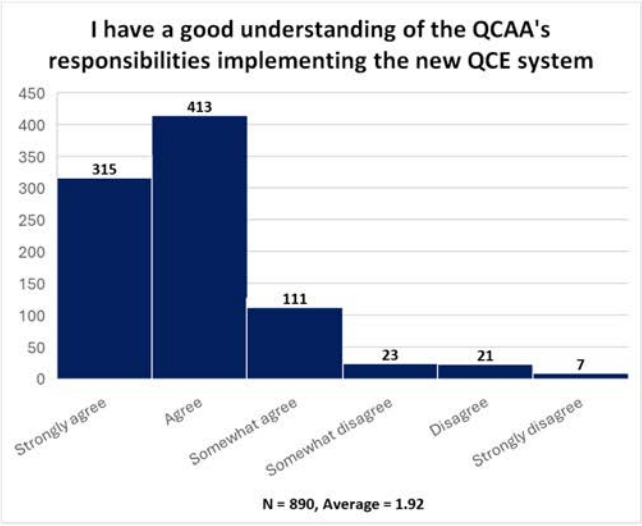
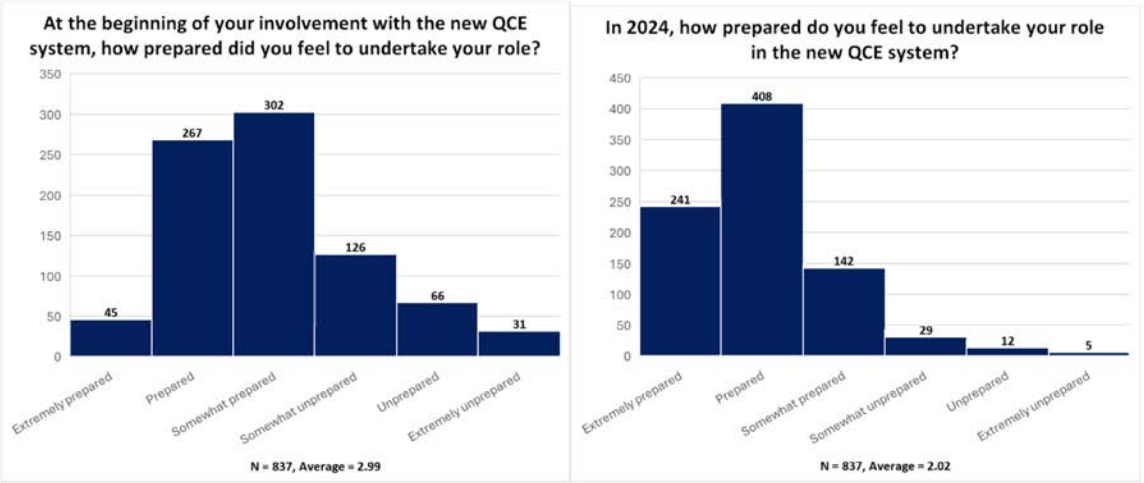


Figure 9.23 Endorsers' perspectives of sustainability and improvement



Chapter 5 Findings: New assessment processes

Survey evidence: Endorsement

The four panels in Figure 9.24 highlight the positive views of the process from the targeted Endorser survey. For the numeric summaries, the categories are coded from 1 (strongly agree) to 6 (strongly disagree), so lower average values indicate more positive responses. In addition to the quantitative survey data, 175 free-response comments were provided on the survey. The most common comments (approximately 14%) indicated that the current Endorsement process is much better than the previous process in the former QCE system, and that it has improved over the years since initiation.

Despite generally positive survey responses regarding the process and timing of Endorsement, shown in Figure 9.25, these were also among the most commented-on themes in the free responses. Just under 10% of the comments addressed the workload and scheduling of the Endorsement events, with concerns about weekend work, the concentration of effort in a very short period, and potential effects of fatigue on judgements. Although the overall rating is not negative and the majority of responses are in 'agree' categories, the survey question 'The Endorsement process is sustainable from a workforce perspective (i.e., recruiting and retaining teachers as Endorsers) in the second panel of Figure 9.26 was the lowest rated across all Endorsement items, reflecting the concerns raised in the free-response comments.

Overall, survey respondents felt very positively about the Endorser training and support they received from their leadership. About 5% of the comments requested more samples and materials to clarify the task requirements, with the goal of increasing their confidence and competence in Endorsement and writing of directives to schools. This request is a positive sign of the investment in producing high quality results for all schools felt by these Endorsers. A similar number of comments suggested consideration of a bank of pre-endorsed assessments that schools can use, and a bank of plain-language Endorsement comments that Endorsers could choose and apply when an instrument is not endorsed. Commenters believed that this would improve quality and streamline the Endorsers' workload.

Figure 9.24 Views on Endorsement process: Quality

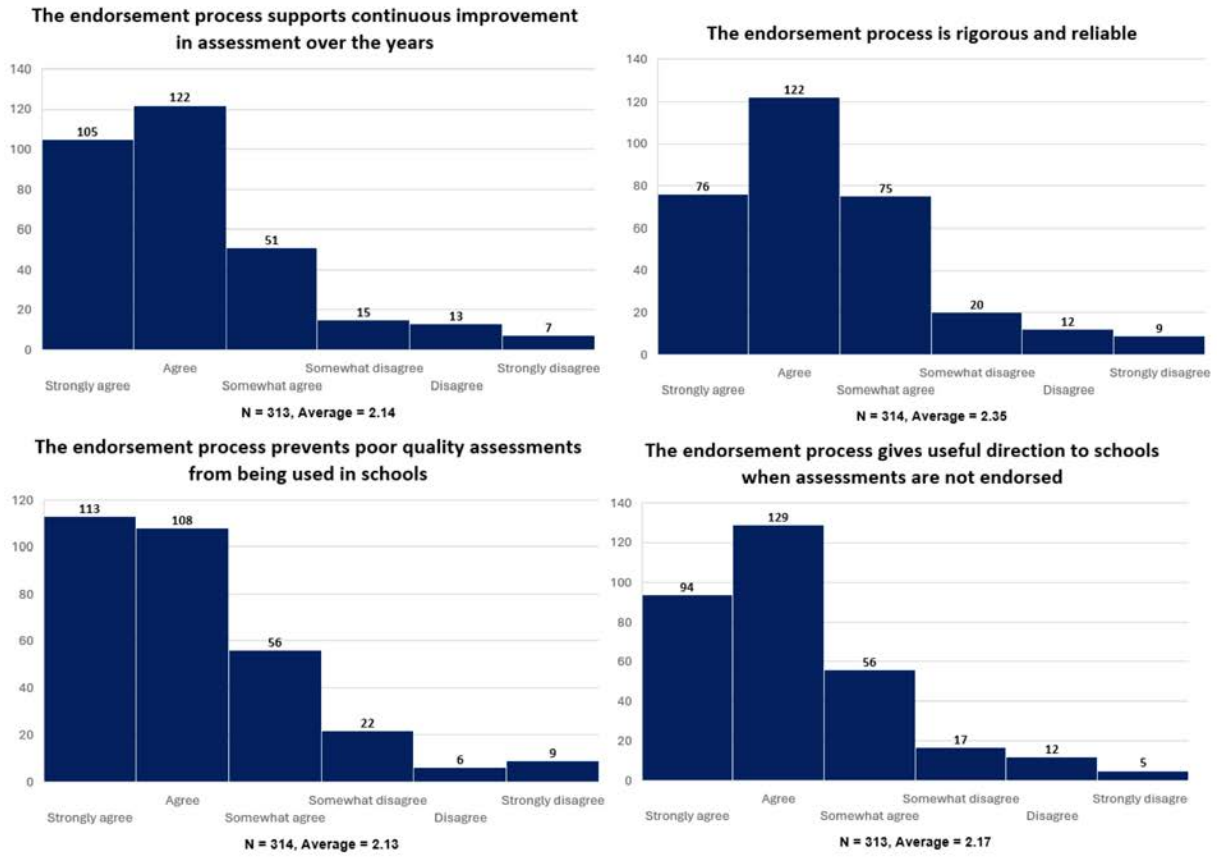


Figure 9.25 Views on Endorsement process: Structure

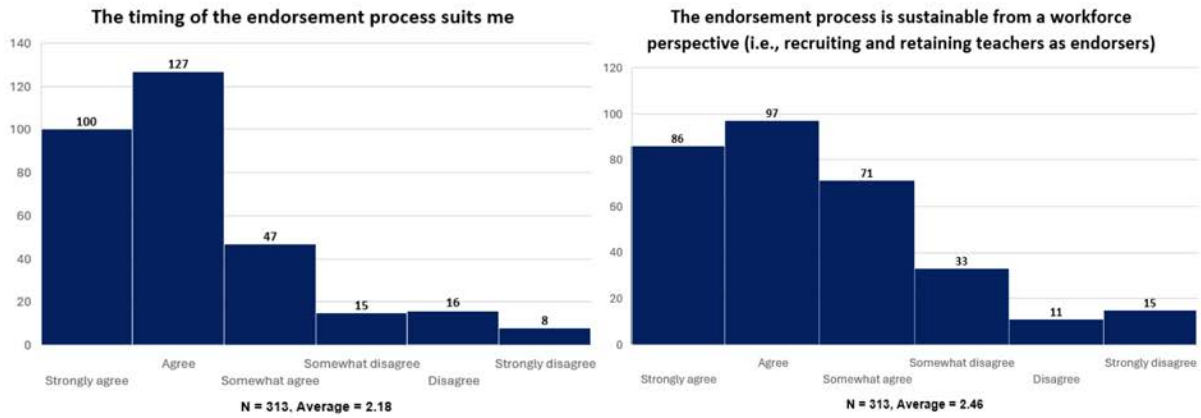
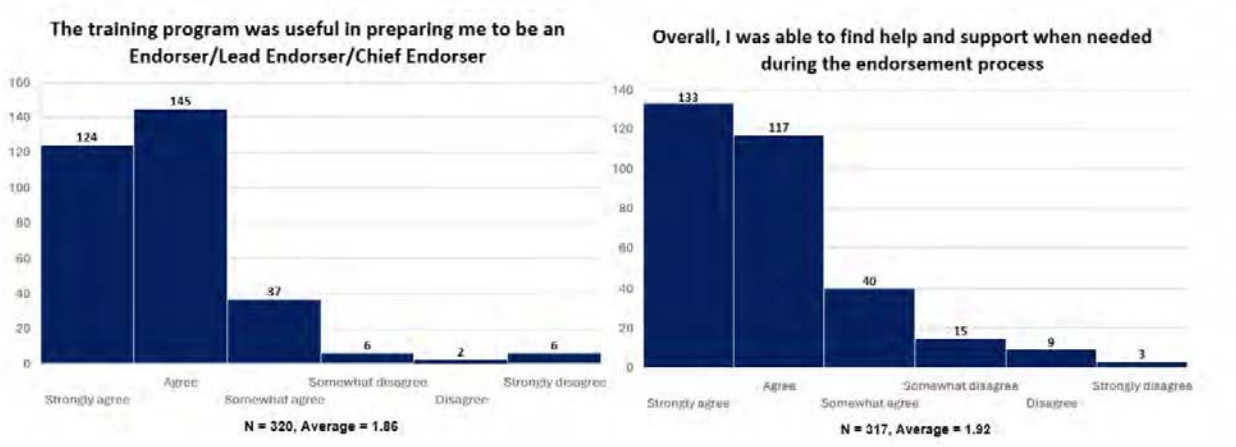


Figure 9.26 Views on Endorsement: Training and support



Survey evidence: Confirmation

Confirmers, Lead Confirmers and Chief Confirmers were asked to respond to a set of survey questions with categorical responses from 'Strongly Agree' to 'Strongly Disagree'. About 600 Confirmers responded to most survey items, with fewer responses from leads and chiefs expected as there are fewer assessors in the leadership roles. Across the board, the survey responses expressed positive views across items on a range of topics.

In addition to the quantitative survey responses summarised below, 284 free-response comments were provided on the survey. In contrast to the Endorsement comments, the Confirmation comments generally were slightly less positive and focused more on specific concerns, including IT issues, timing and workload, sufficient numbers of available Confirmers and a perceived lack of feedback on Confirmer performance and decisions.

The four panels in Figure 9.27 highlight the positive views of the process from the targeted Confirmer survey. For the numeric summaries, the categories are coded from 1 (strongly agree) to 6 (strongly disagree), so lower average values indicate more positive responses. The item 'The confirmation process is an essential quality assurance process in the QCE system' in the first panel of Figure 9.27 shows particularly positive views on the essential role that Confirmation plays in the QCE system.

As with Endorsement, the survey responses were generally positive about the process and timing of Confirmation, shown in Figure 9.28. The 'timing' ratings were slightly higher (i.e., more negative) than those given to Endorsement. The overall rating is positive, and most responses are in 'agree' categories. Again, these were also among the most commented-on themes in the free responses. 10.6% of the comments were around the workload and scheduling of the Confirmation events, with concerns expressed about weekend work, overlap with holidays, insufficient number of Confirmers to complete the work in a timely fashion, the concentration of effort in a very short period and potential effects of fatigue on judgements.

There was increased disagreement among the Confirmers regarding the sufficiency of the allocated time to complete their duties during the event, and this was more evident among the Lead Confirmers. Nearly 30% of responses received from Lead Confirmers were in a 'disagree' category and this item had the highest (i.e., most negative) rating on the Confirmers survey. QCAA may wish to conduct follow-up interviews with the Lead Confirmers on this topic. The responses may indicate that Lead Confirmer review is being completed as required to complete the outcome decision, but the review may not be as detailed or thorough as desirable due to time constraints.

The level of confidence in the accuracy of Confirmation decisions increases steadily across the leadership hierarchy, with leads and chiefs expressing greater confidence as seen in the survey responses shown in Figure 9.30. This may be interpreted as an effect of more experience and training for those in leadership roles.

Approximately 7.4% of the free-response comments noted a lack of feedback to the Confirmers from leadership and a sense of not knowing whether the Confirmer's changed marks and their key cited evidence persisted through to the final stage or was revised/not used by the Lead or Chief Confirmer in delivering the school results.

Nearly 15% of the free-response comments on the Confirmer survey were about technology, focused on the Confirmation app, with multiple mentions of server overload and extended lag and load-times in the app. It seems that there was an IT issue close in time to the survey administration, as several of the comments referenced 'today', so the frequency of this category of comment is likely higher than typical. Confirmation is undertaken remotely, and it may that local internet service contributed to the issues encountered, especially in high-bandwidth subjects with video, media or large images to be reviewed. There were also comments on the app layout, image resolution and browser incompatibility. Despite frequent comments on IT issues and concerns, however, the ratings of the Confirmation app were positive.

Figure 9.27 Views on Confirmation process: Quality

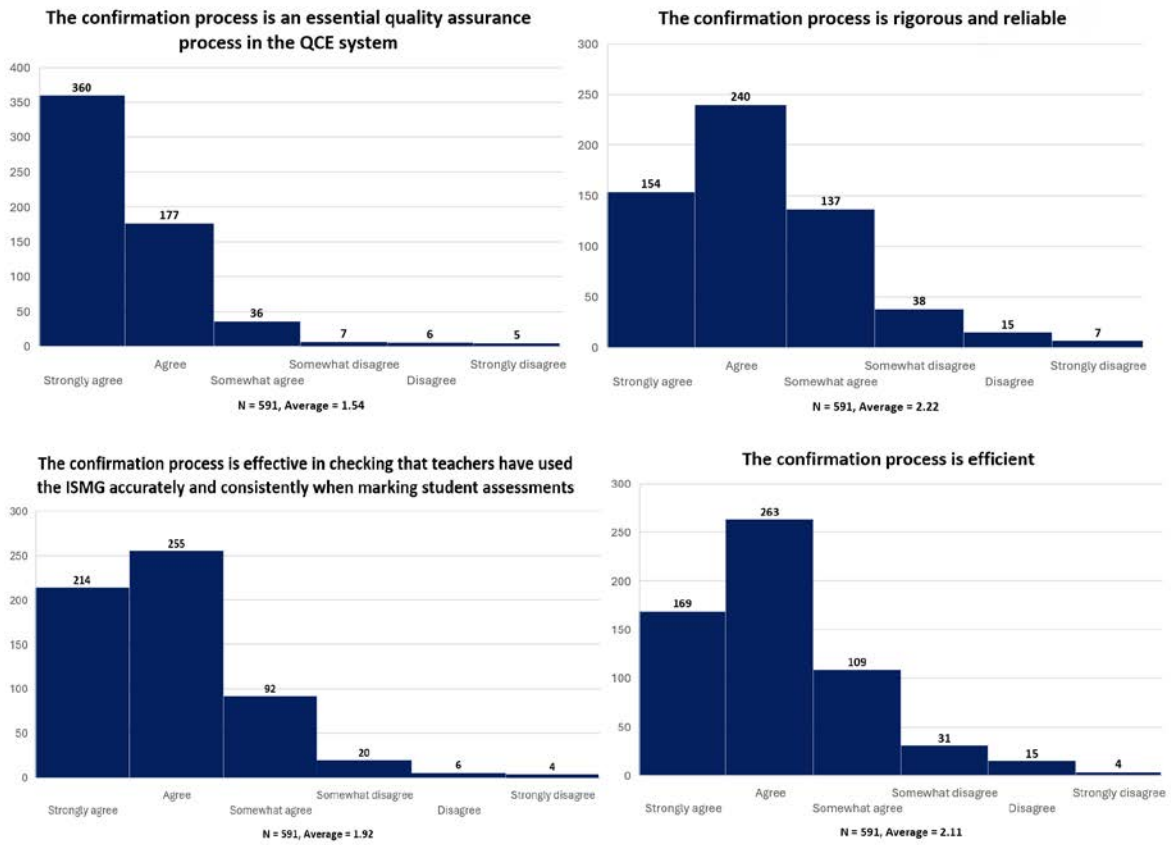


Figure 9.28 Views on Confirmation process: Structure

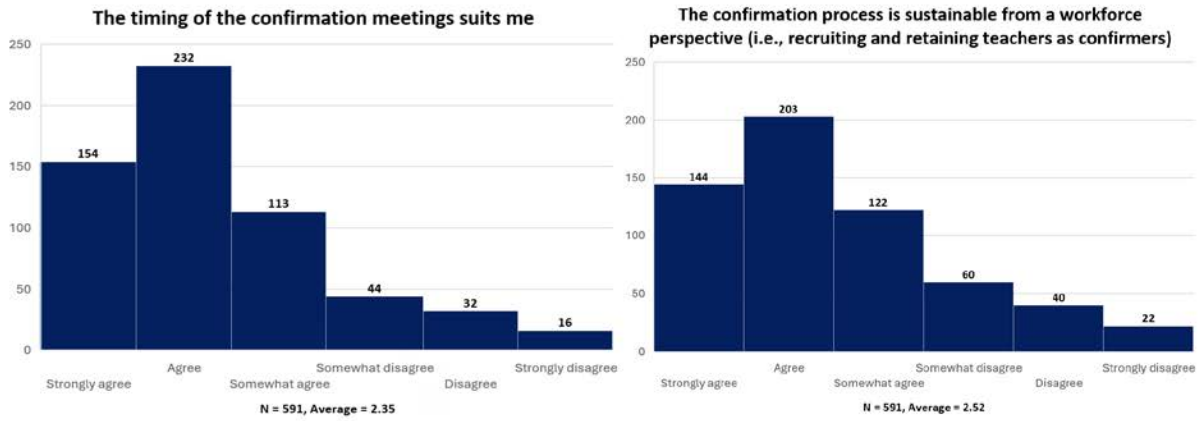


Figure 9.29 Views on Confirmation process: Duties and workload

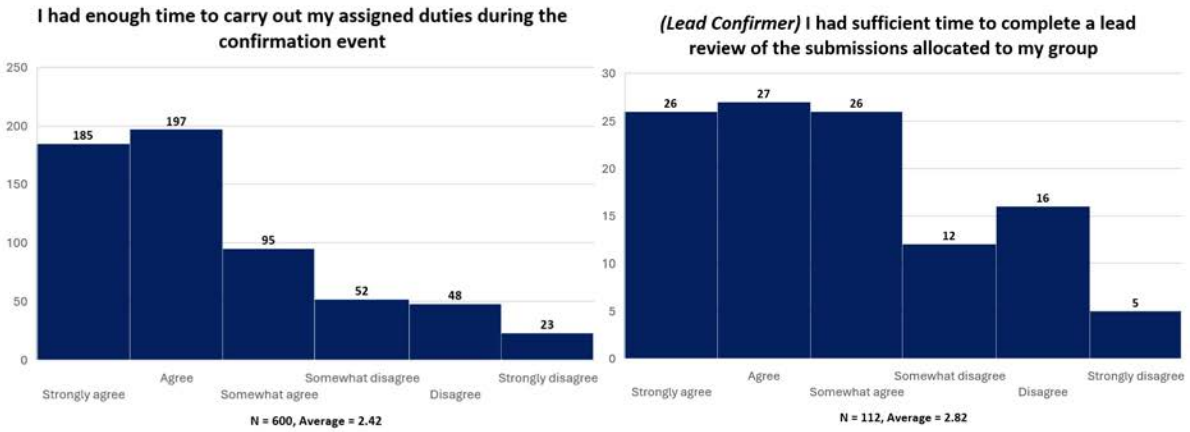
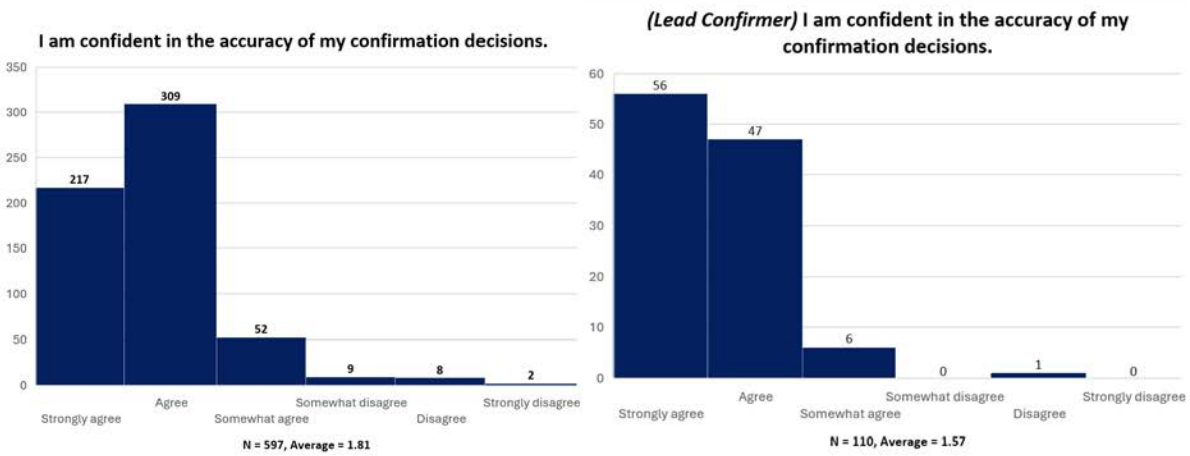


Figure 9.30 Views on Confirmation process: Confidence in decisions



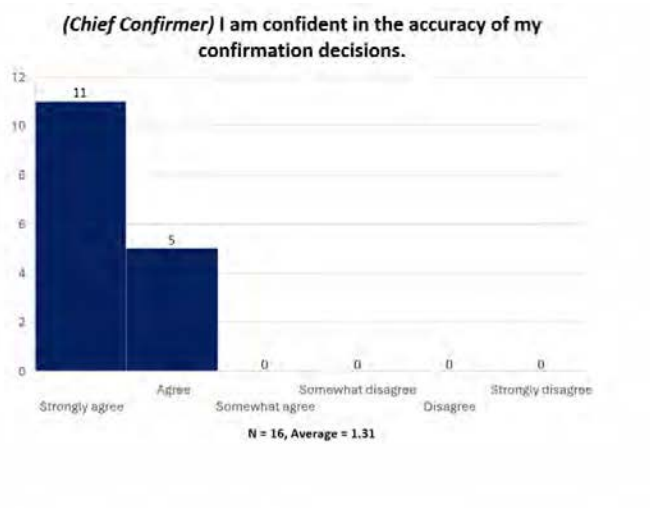
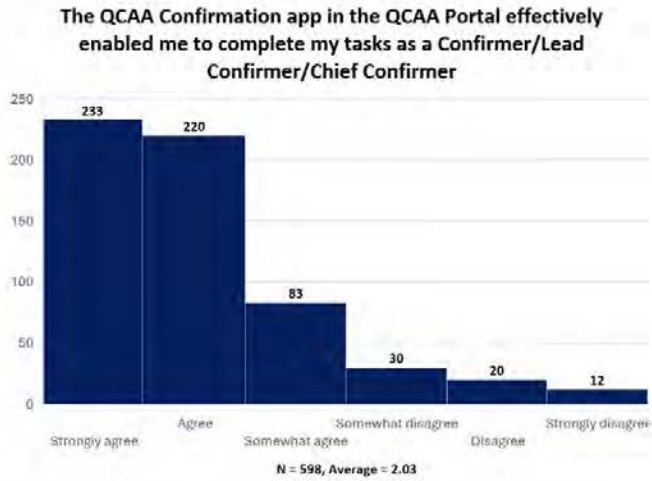


Figure 9.31 Views on Confirmation: Technology



External assessment: Markers and stakeholders

The External Assessment (EA) Markers were surveyed to gain qualitative data on their experience. For the numeric summaries, the categories are coded from 1 (strongly agree) to 6 (strongly disagree), so lower average values indicate more positive responses. In addition to the quantitative survey data, 498 free-response comments were provided on the survey. The EA Markers held positive views of the assessments, marking process and security measures, as can be seen in Figure 9.32.

As with Endorsement and Confirmation, the survey responses were generally positive about the process and timing of marking, shown in Figure 9.33. One observation: in the 'timing of reassessment period' item in the first panel, the number of responses is lower than for the other items in the survey. This may be attributed to the fact that most Markers are not involved in reassessment; it may also be that Markers misinterpreted the item as referring to the EA marking period rather than reassessment and responded on that basis.

Among the most commented-on themes in the free responses were workload and pay rates. 5.6% of the comments received were around high workload and scheduling of marking, with concerns expressed about weekend work, overlap with school activities and assessments at the end of term, the expectation of 4 hours marking after a full day working, pay based on piecework which resulted in low rates, and potential effects of fatigue on judgements. The concern about the 4-hour marking period after a full working day may be based on a misunderstanding of QCAA's stated requirements, which are that markers must be contactable for a 2-hour period between 4 and 8 pm and that the minimum marking session be 2 hours.

In contrast to the survey results for the Confirmers, the Markers felt that the expected pace and period of their work was achievable. The Lead Markers, although still giving positive ratings, appeared to feel more time pressure as their rating was higher than the Markers (i.e., more negative), as was their rating on whether they felt they had sufficient time to prepare for marker training.

Markers and the Lead and Chief Markers expressed confidence in the accuracy of their marking. Again, there was a pattern of increased confidence across leadership roles, likely paralleling additional training and experience in the process. And despite the expressed confidence in marking accuracy, the most negatively rated item in the Marker survey was the question asking whether there was useful data to track marker performance. Feedback was the most commented theme in the free responses, with 7.4% of the comments requesting more feedback sooner in the process. This concern echoes that expressed by the Confirmers, also seeking more concrete and useful feedback to improve performance.

Stakeholders who participated in the EA marking operations had positive feedback about the event, noting that it was well-organised and efficient. This finding is consistent with the observations conducted by the evaluation team. Observations of script selection, marker training, online marking, quality assurance, logistics and the Illness and Misadventure (IM) committee revealed a thorough system focused on ensuring high-quality, consistent approaches in all aspects. EA operations were seen as focused on fairness to the student above all. Script selection, marker training and online marking were supported with detailed templates and procedures that resulted in consistent approaches taken across subjects, while still allowing for variations due to the nature of the subject.

Figure 9.32 Views on Marking process: Security

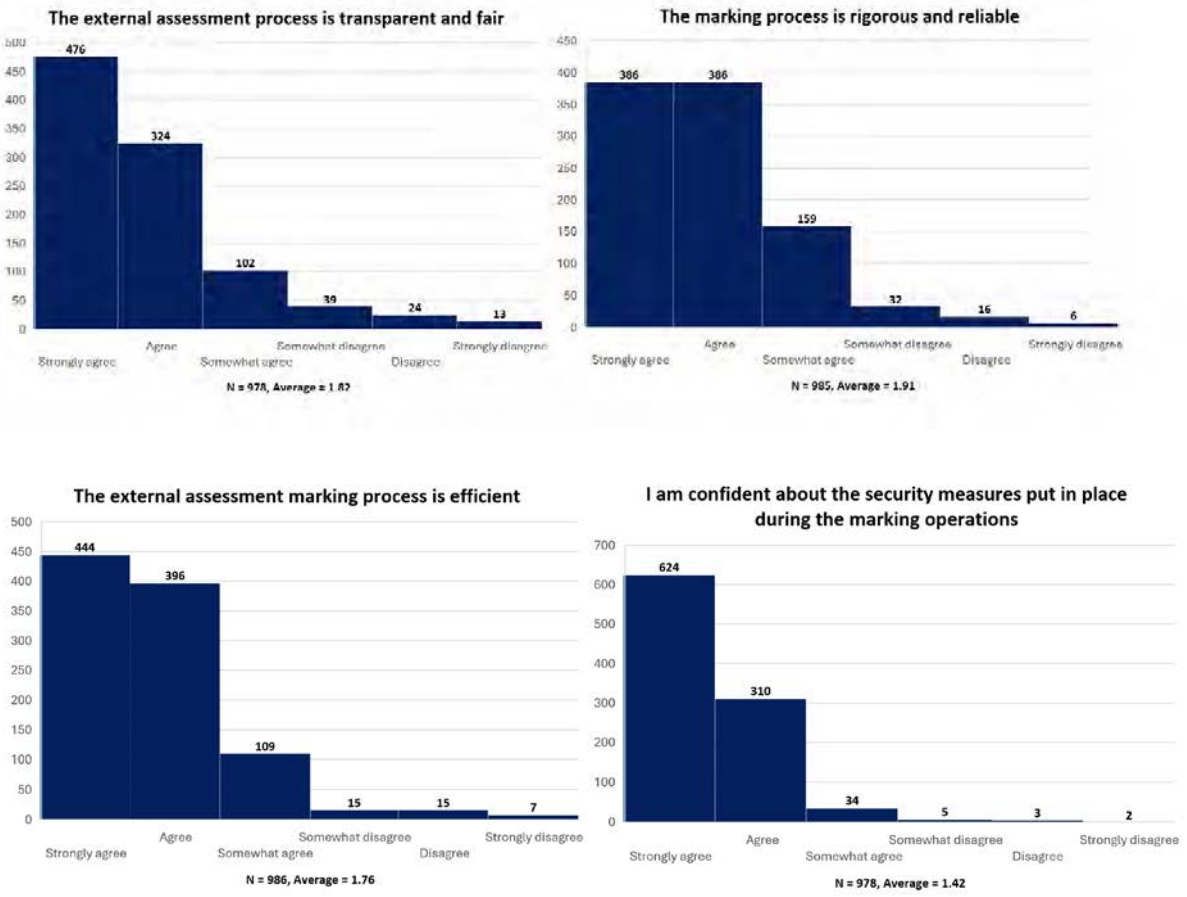


Figure 9.33 Views on Marking process: Structure

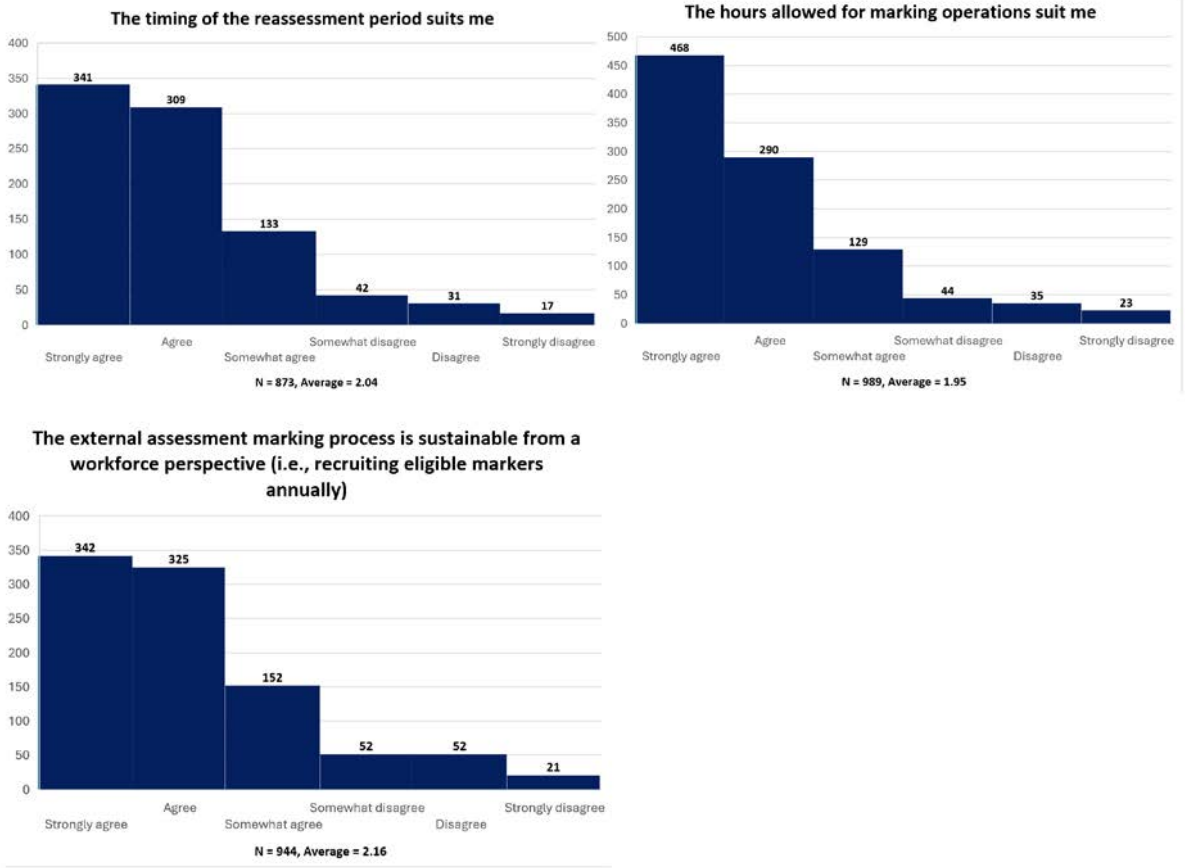


Figure 9.34 Views on Marking process: Duties and workload

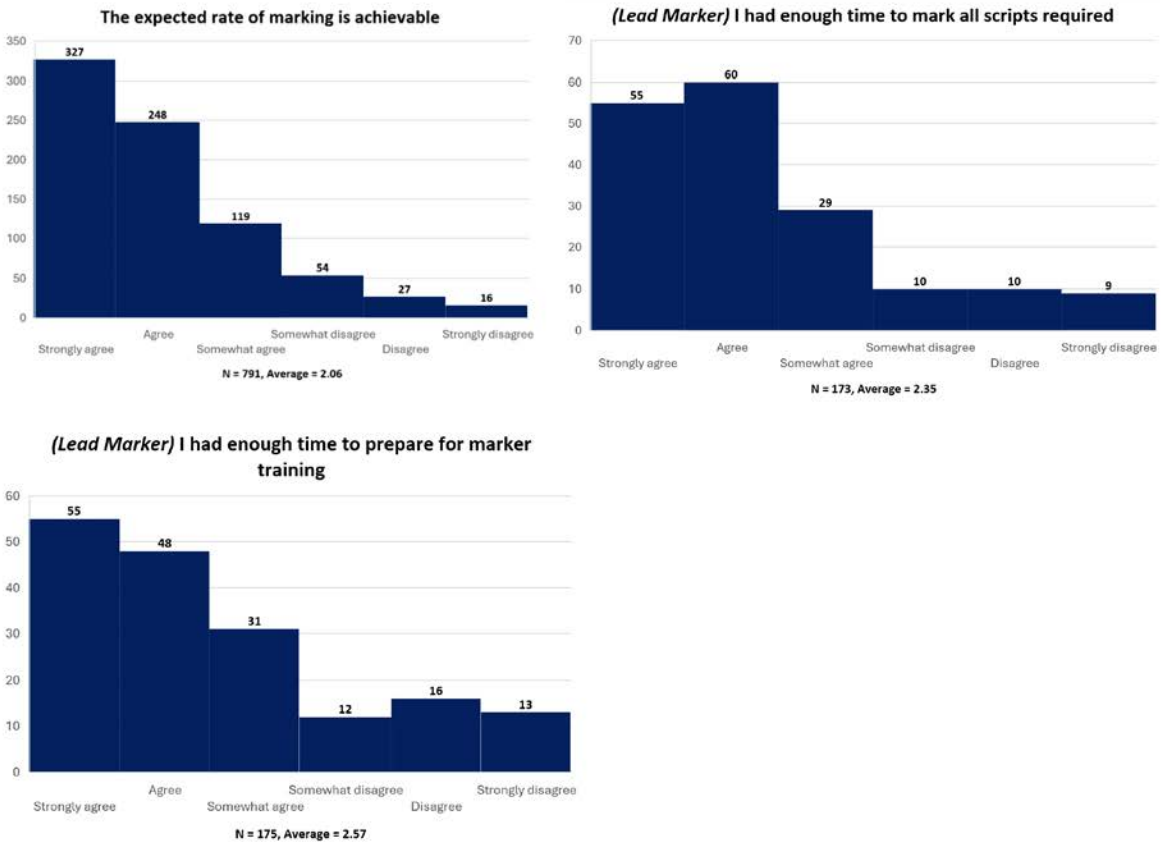
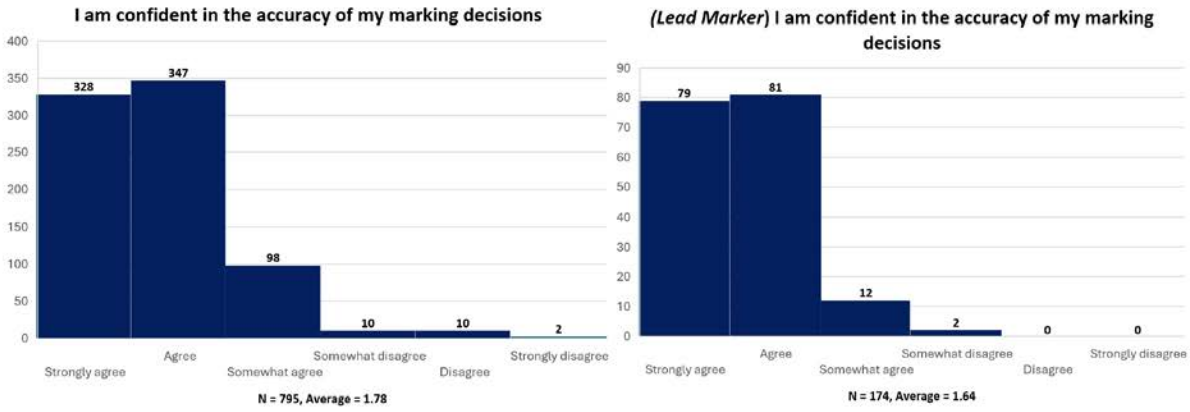
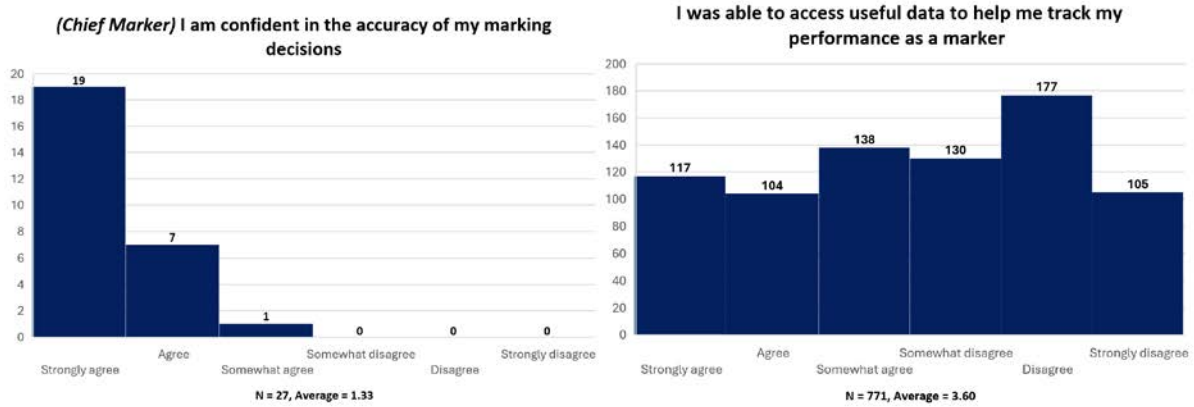


Figure 9.35 Views on Marking process: Confidence in decisions





Questions were raised in some stakeholder interviews and focus groups regarding the sustainability of the casual workforce model, especially its dependence on teachers being available through release from instructional days, or on evenings, weekends, or holidays. These concerns were also raised by surveyed QCAA staff.

QCAA staff views

QCAA staff were also surveyed and provided free-response commentary. Their surveys were not targeted to specific phases of the quality assurance process, and so are presented in this section with a focus on topics that echo those raised in the targeted surveys.

QCAA staff reiterated the confidence expressed in other surveys, supporting the new system and the quality of the training and support provided to schools and students. The comments were similarly positive, with 15.6% noting increased fairness and equity in the Internal and External Assessments and 11% commenting on the clarity and quality of the syllabus and curriculum documents.

Staff survey results and comments also reflected concerns about the workload and sustainability of the new process, especially Confirmation. Approximately 23% of the respondent comments acknowledged the quality and consistency Confirmation has brought to the QCE assessment system but expressed concerns about an unsustainable workload and potential attrition of the necessary casual workforce. Some concerns were raised by QCAA staff regarding potential conflict of interests for assessment writers involved in writing textbooks and developing other subject resources.

Figure 9.36 QCAA staff views: Quality of new QCE system and support

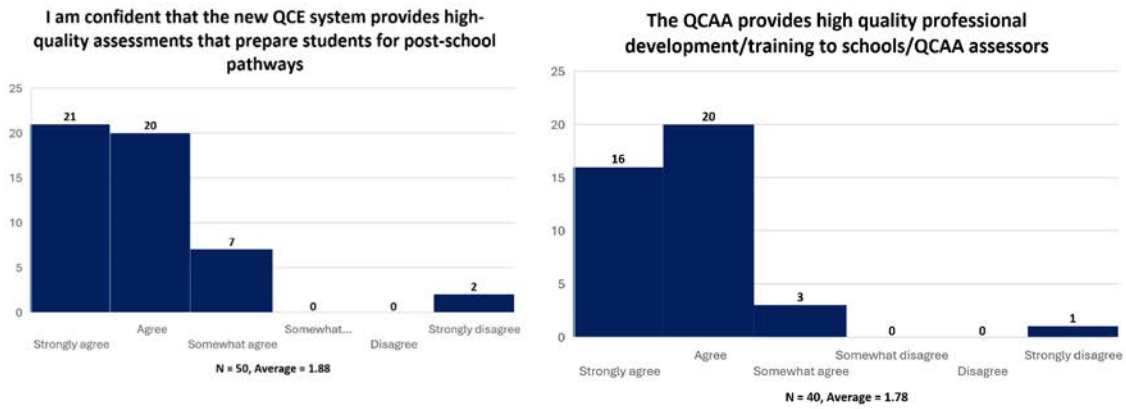
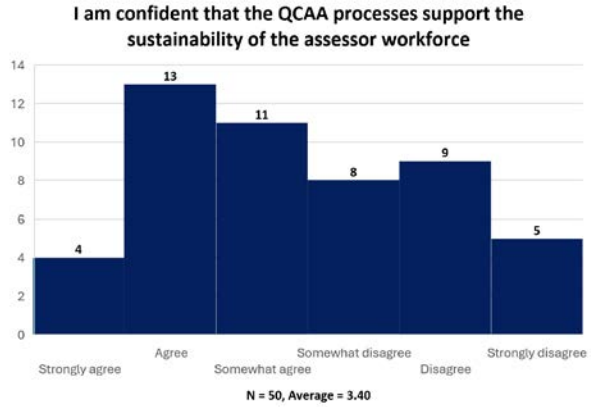


Figure 9.37 QCAA staff views: Sustainability



Stakeholder survey responses

Stakeholders were also surveyed as part of this evaluation. The targeted groups included school staff members, past QCE students, parent or carer of a current or former QCE students, and other. The 'other' category included, for example, academics, tutors and retired educators. The survey branched depending on the reported role of the respondent, so these survey items will have varying numbers of responses.

Overall, External Assessment was viewed positively by stakeholders, as can be seen from the responses in Figure 9.38. The EA was the most commented-upon aspect when asked 'What works well in the new QCE system?', with 15.3% of 876 comments. The next most common theme in the comments (12.8%) was around consistency in the assessment construction and marking, attributed largely to the availability of high-quality common syllabuses. The themes of the comments covered operational delivery, equity, fairness, academic rigour, explicit content and instructions, and high-quality marking procedures.

Resources and support provided by the QCAA were also viewed positively as supporting teachers and schools to successfully navigate the system requirements.

Stakeholder views on the system as whole were mixed. Some specific patterns observed include:

- The most commented topic when asked 'What does not work well in the new QCE system?' was the weighting of Internal versus External Assessments across subject areas. 18.9% of the 387 comments were on this topic. In addition, weighting and scaling received 12.0% of the 599 comments from the item requesting any other feedback. While there was no consensus on whether External Assessments should be weighted at 25% or 50%, participants strongly expressed the view that all subjects should have the same EA contribution.
- Some comments noted that the scaling system was perceived to be unfairly biased against arts content, pushing students to STEM subjects that they may not enjoy. Note that inter-subject scaling is not a responsibility of QCAA but instead resides with QTAC. Clarification of realms of responsibility with stakeholders may direct these concerns to the appropriate agency.
- Additional themes noted in the comments from stakeholders were:
  - pressure on teachers and students
  - a perceived drive towards 100% QCE attainment
  - a crowded syllabus that does not permit sufficient time for learning all the content

The above concerns formed 6.7% of the 387 'What does not work well in the new QCE system?" comments and 7% of the 599 comments on the 'Any other marking feedback' item. Again, note that QCAA does not promote or require 100% student QCE attainment. This pressure is likely coming from other sources such as schools, systems and their governing bodies.

These concerns likely contributed to the slightly higher (i.e., more negative) ratings when compared to views of other groups such as the assessors or QCAA staff on the QCE system survey items shown in Figure 9.40.

Interestingly, stakeholders saw implementation of the QCE system through two different lenses, depending on whether the context was their own school or other schools, as shown in Figure 9.40. They had positive views of the implementation in their own school, but expressed concerns about consistency across other schools, sectors and contexts. This phenomenon of rating local schools highly and schools in other contexts such as statewide or nationally less so is not unique, as West (2014) and the Phi Delta Kappan annual poll (PDK, 2022) show.

Figure 9.38 Stakeholders' views: Assessment

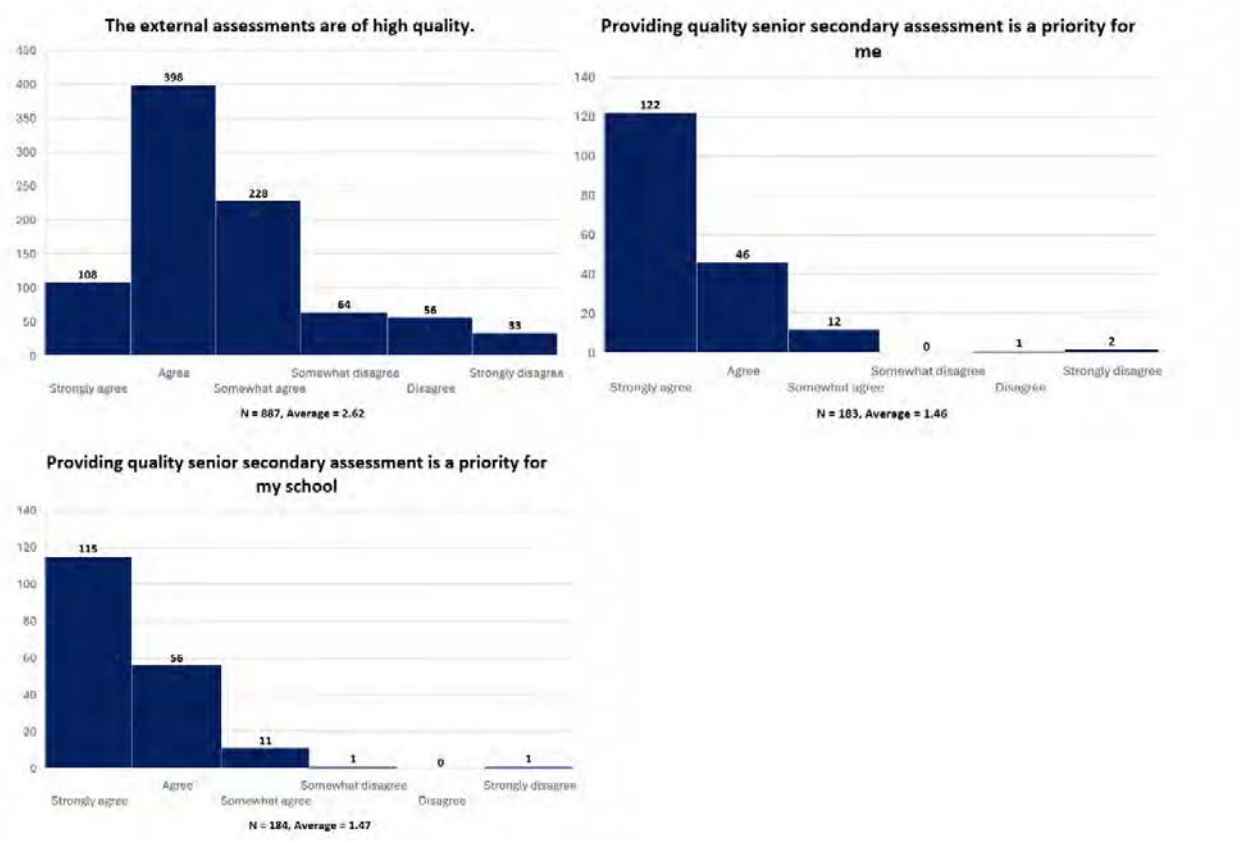


Figure 9.39 Stakeholders' views: QCAA resources

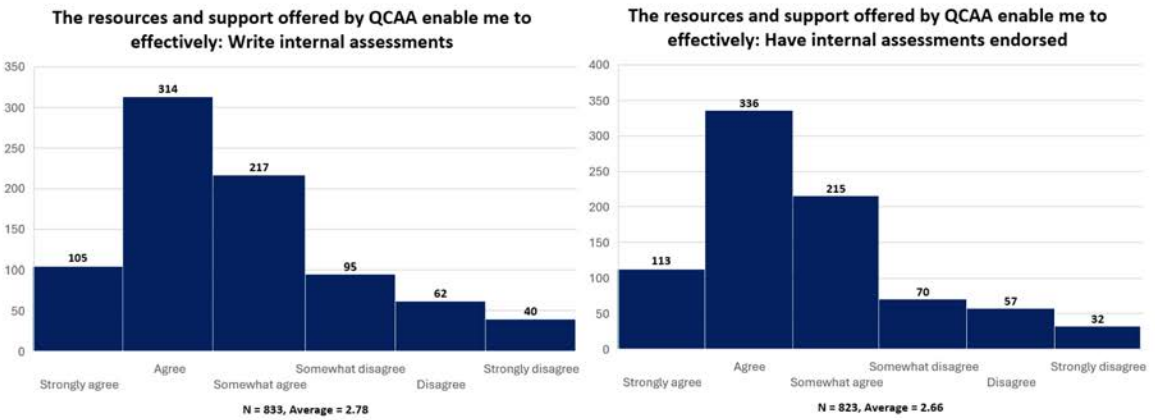
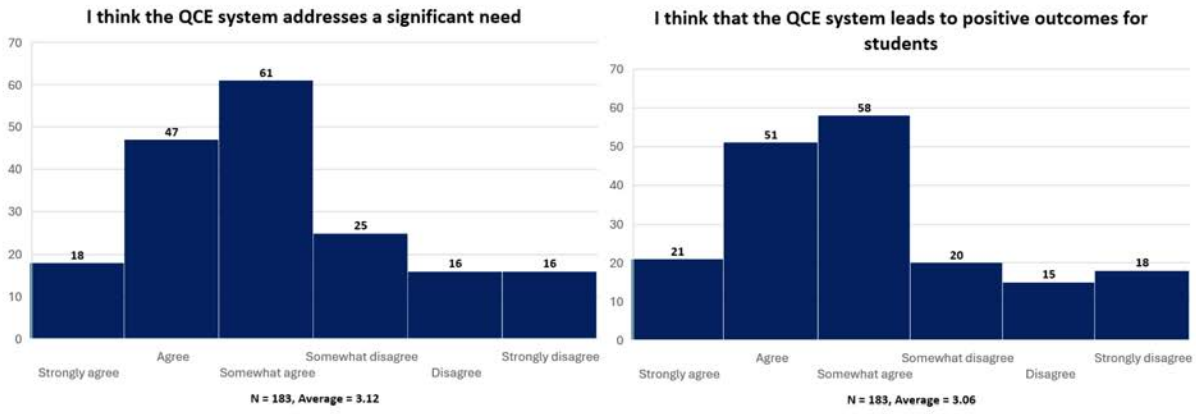
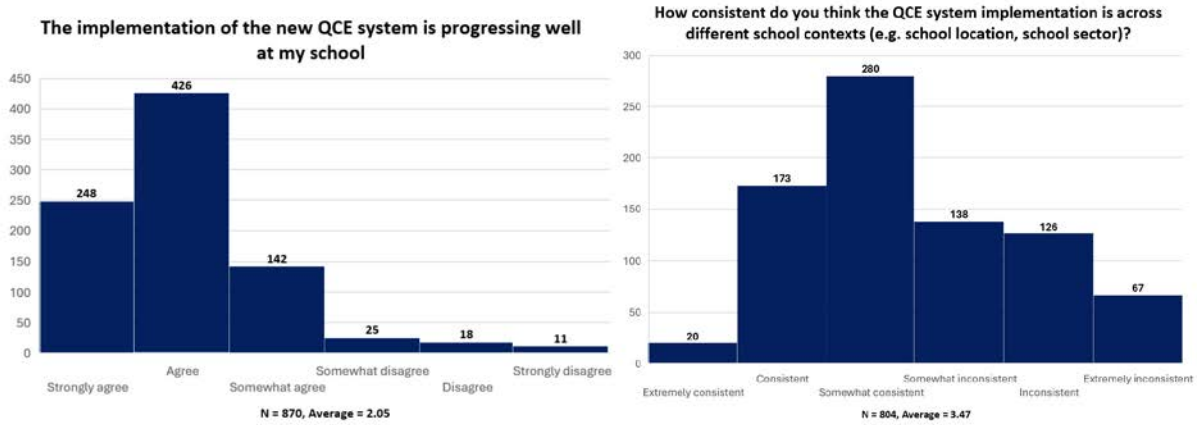


Figure 9.40 Stakeholders' views: QCE system and implementation

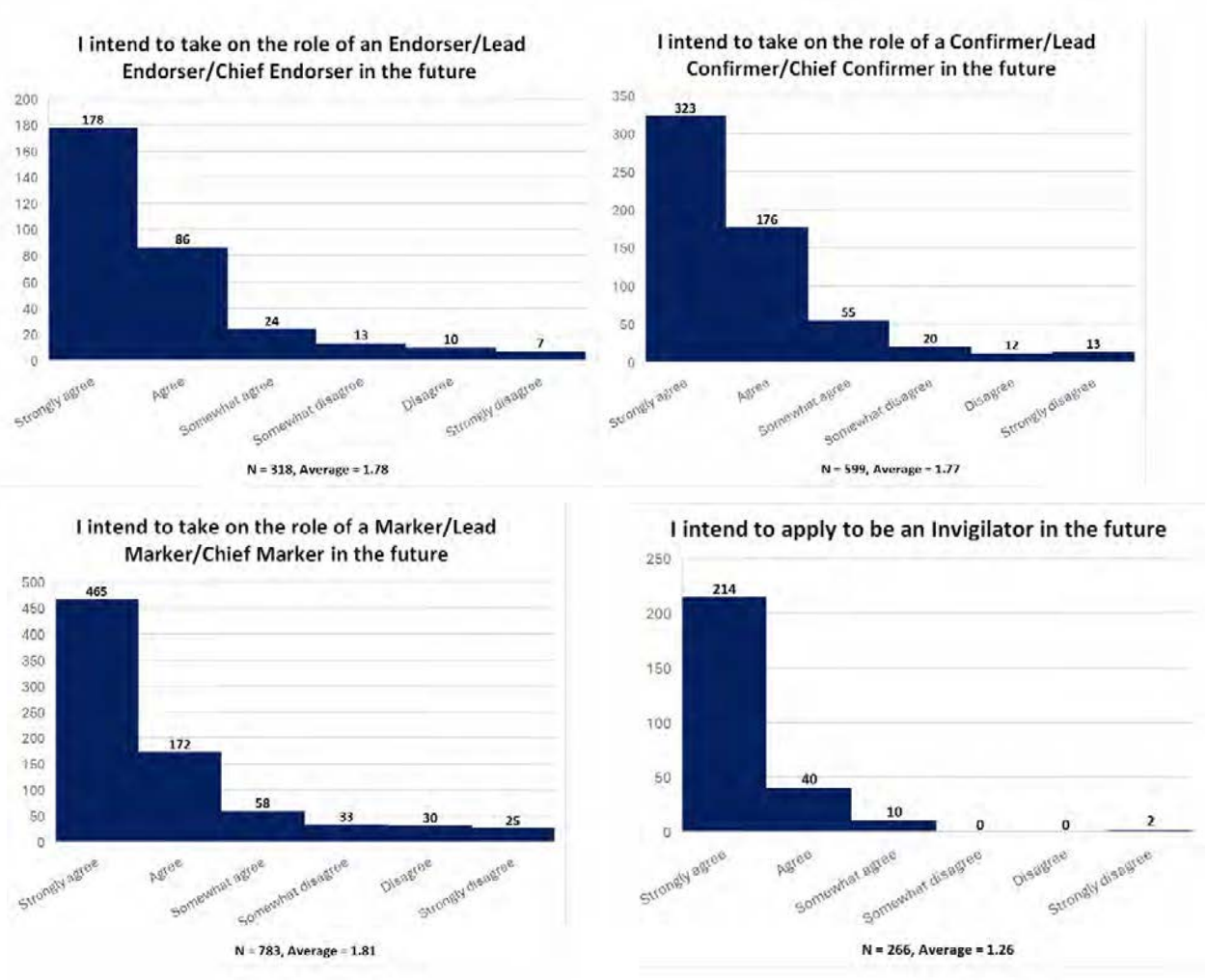




Effectiveness of casual assessor recruitment and retention strategy

Survey data from those currently employed in casual assessor positions indicated that many would take on the role again in the future. For example, 83% of those in the Endorser survey agreed that they would take on the role in the future (see Figure 9.41). Such strong agreement is a good indicator that teachers enjoy these roles and that the positions fit well within their existing work requirements.

Figure 9.41 Casual assessors' interest in future roles

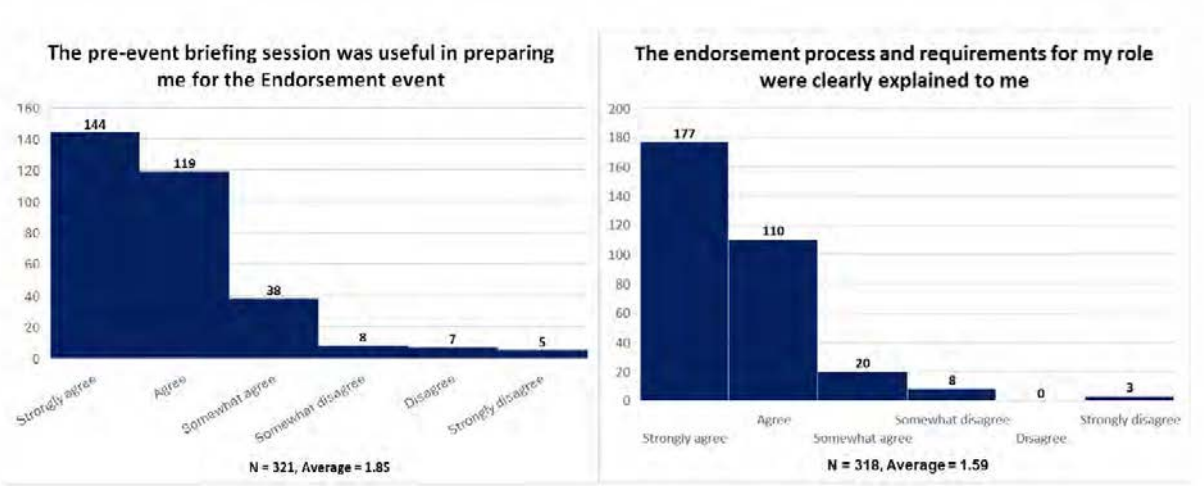


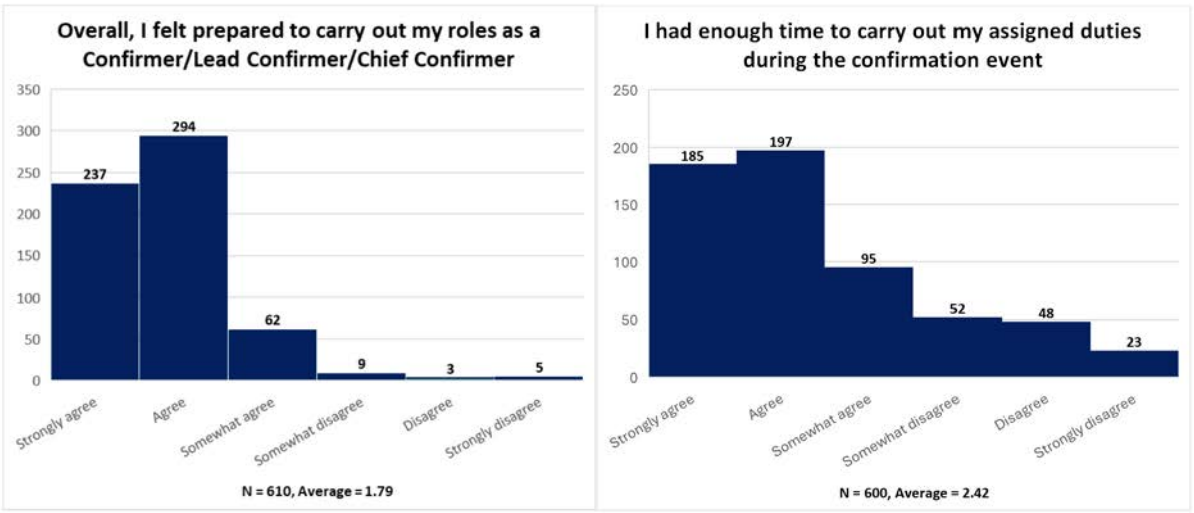
This is supported in data for questions asking casual assessors about the details of their roles, from the application process to the workload expectations. Overall, casual assessors felt that they were well prepared for their roles and that they had the appropriate workload to be able to complete their duties. Details for each of the four categories of casual assessors are detailed below.

Endorsers

Endorsers tended to agree that pre-event training helped them to prepare for the Endorsement events and that they were generally prepared for their roles. They agreed that training opportunities (e.g., online modules, training programs) and resources (e.g., assessment and evaluation tools, Endorsement app, feedback and support from Lead and Chief Endorsers) helped them to feel prepared to complete their work and to carry out their role. They also tended to agree that they had enough time to complete their role requirements.

Figure 9.42 Endorsers' reports of role details

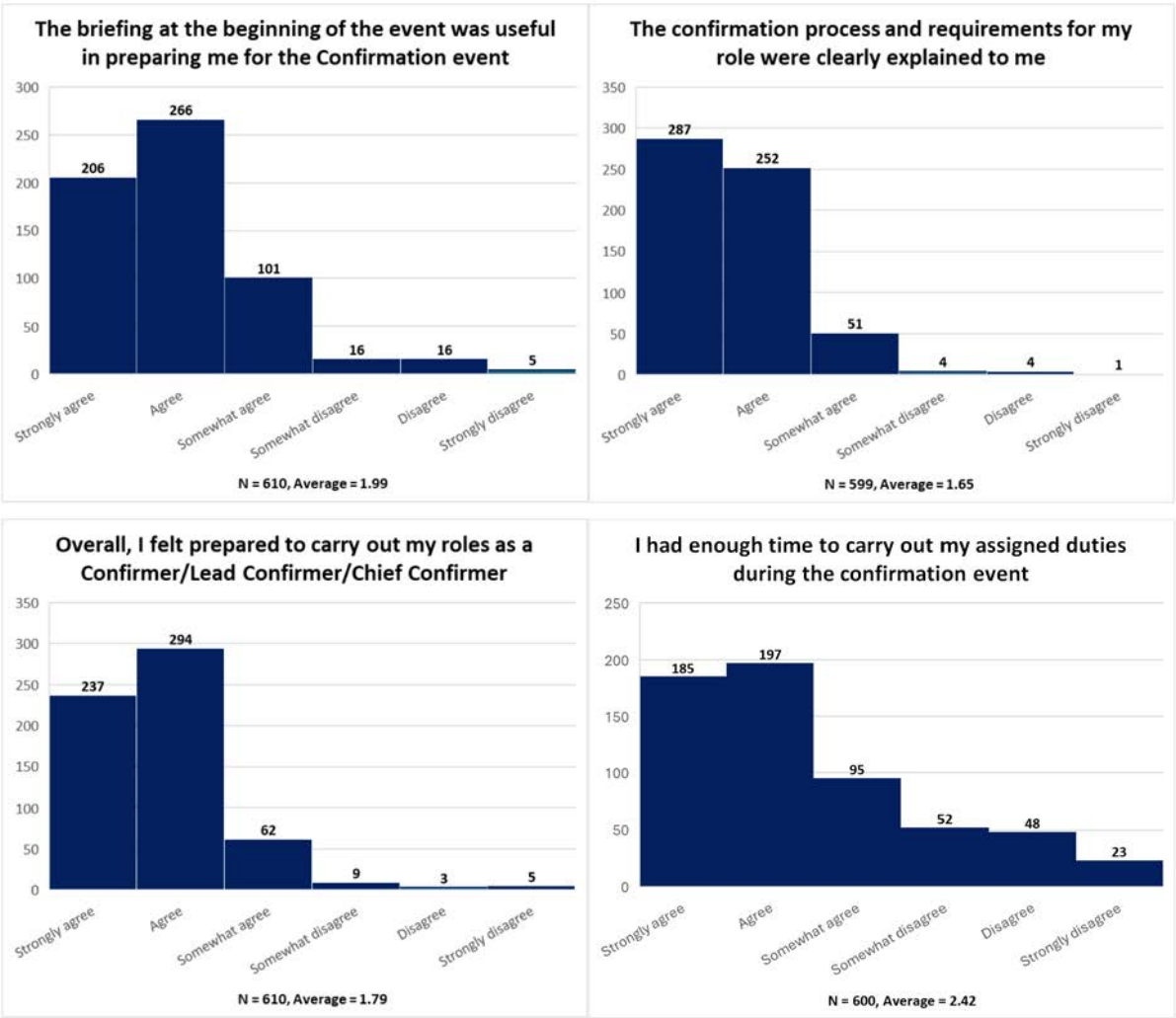




Confirmers

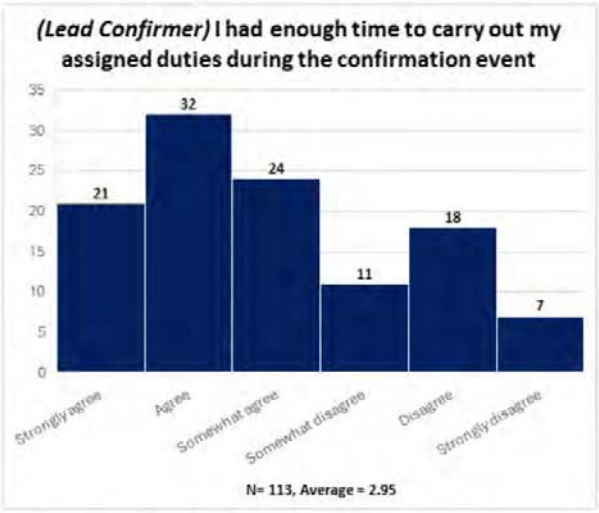
Similar patterns were found for those in Confirmer roles. Confirmers tended to agree that they had adequate training, resources, and time to complete their roles efficiently and effectively. There were some slight deviations for Lead Confirmers in their responses to some questions.

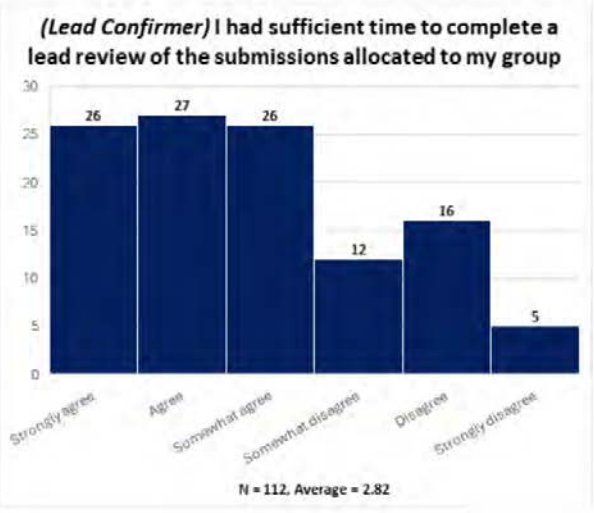
Figure 9.43 Confirmers' reports of role details

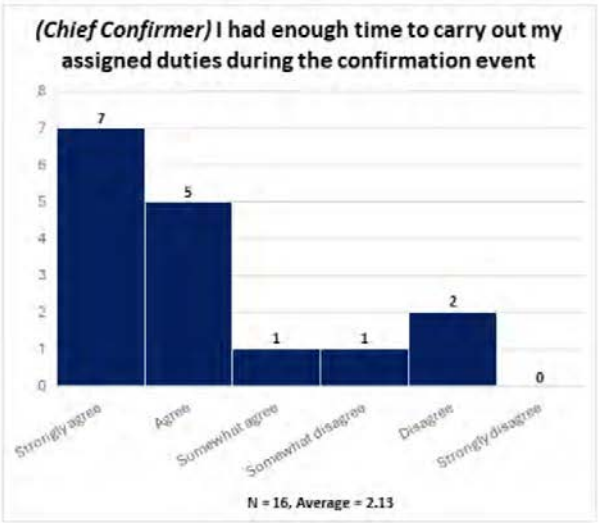


Lead Confirmers were slightly less likely to agree that they had enough time to carry out their duties (47% Lead Confirmers agreed or strongly agreed with this statement vs. 64% of Confirmers). This may be due to the time required for completing a lead review of the submissions allocated to the Lead Confirmers' groups. Chief Confirmers' reports on time allocation were similar to Confirmers.

Figure 9.44 Lead and Chief Confirmers' reports of time allocation



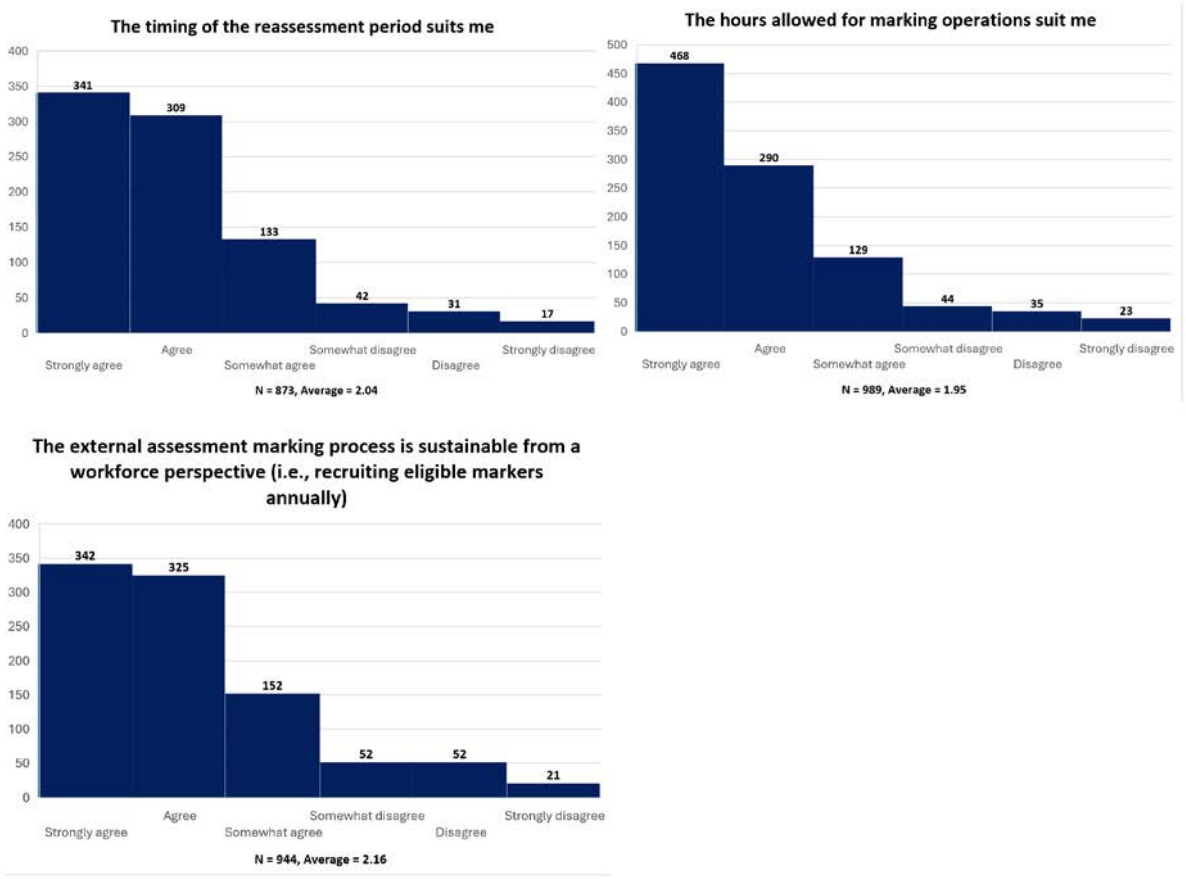




Markers

Those in marking positions positively indicated that the application process was straightforward, that the requirements of the role were clear, they felt prepared for their role, and that the expected rate of marking was achievable. These patterns were similar across those in the Marker, Lead Marker, and Chief Marker positions. Moreover, Lead (74%) and Chief (85%) Markers also tended to agree that their additional responsibilities in recalibrating processes were clear.

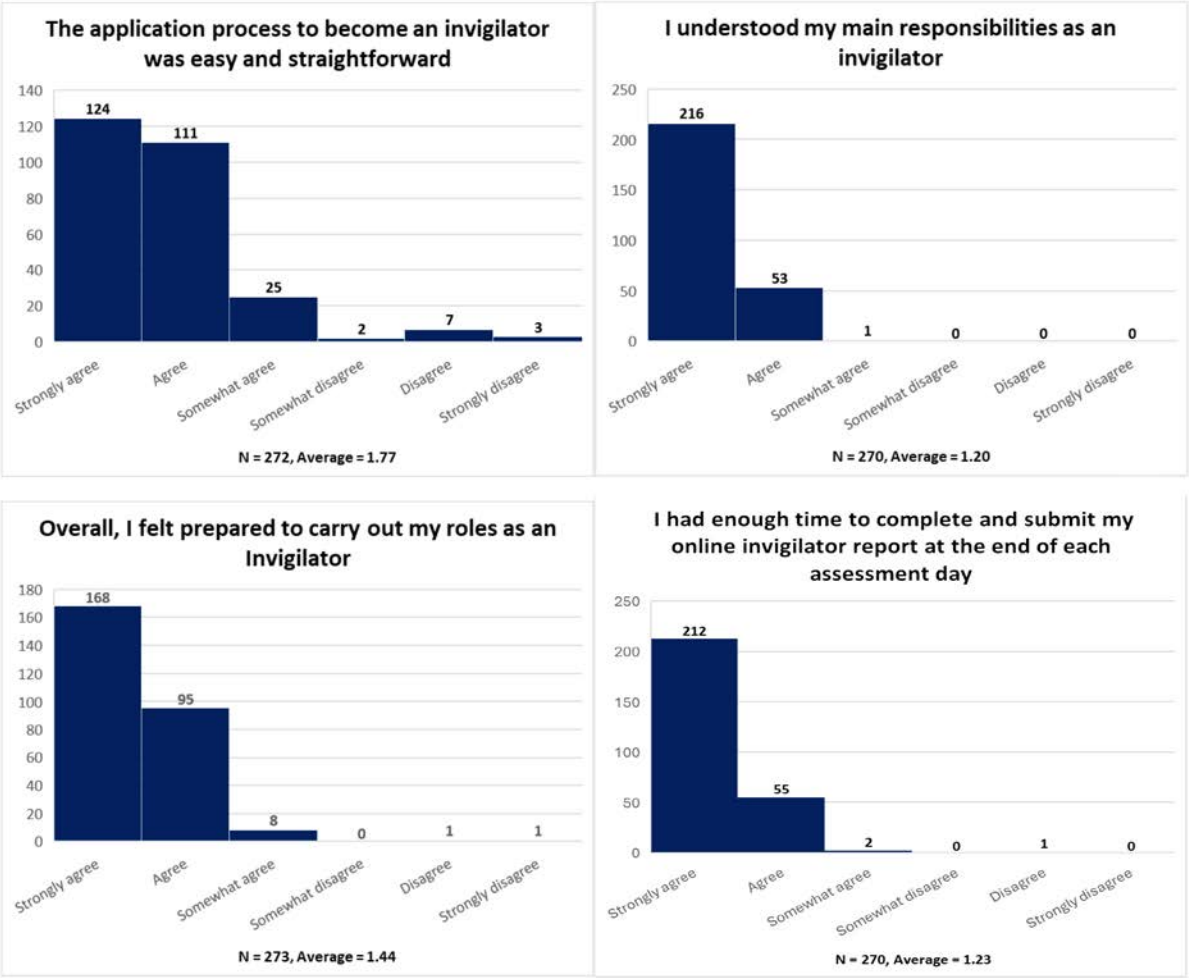
Figure 9.45 Markers' reports of role details



Invigilators

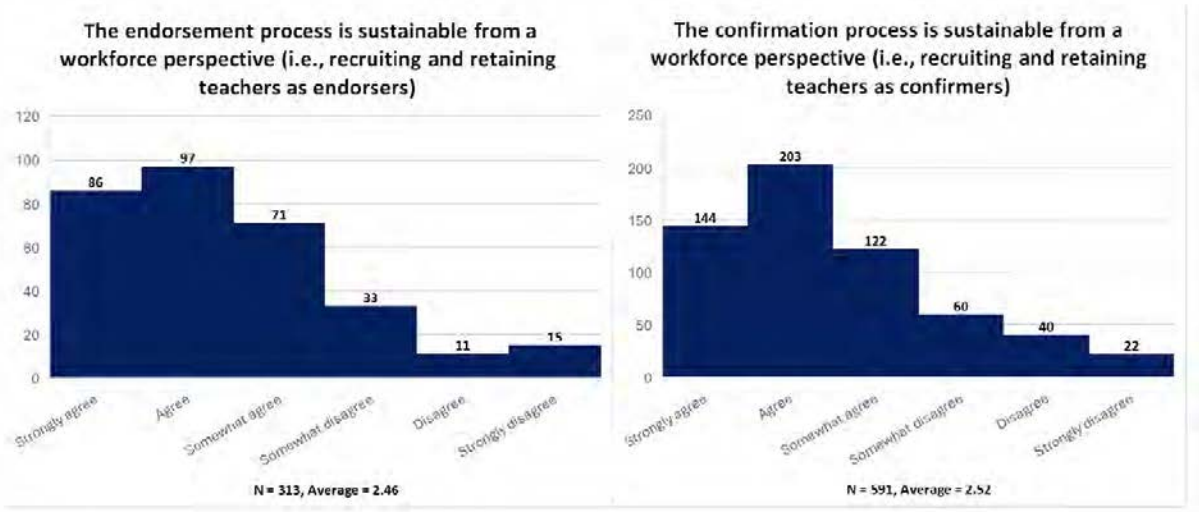
Invigilators' responses followed a similar pattern, with the majority agreeing that the application was straightforward, the role was clear and there was appropriate time allocated to complete the workload Invigilator reports. Most Invigilators also agreed that they felt well prepared for handling issues that arose during assessment events (95%).

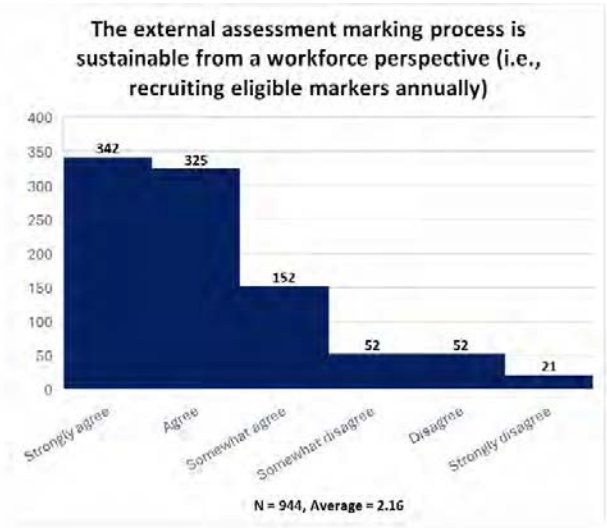
Figure 9.46 Invigilators' reports of role details



Taken together, when considering casual assessor positions within the new QCE system, there appears to be strong satisfaction with the recruitment process and with the duties required for the roles. This is a good indication that the training and support provided to those in these positions will help to retain casual assessors in their roles over time.

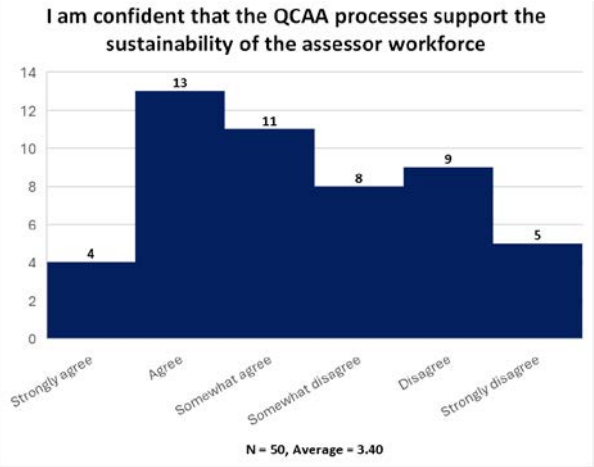
Figure 9.47 Casual assessors' views of the sustainability of their own positions





However, several stakeholders in the QCE system were concerned about the sustainability of the casual assessor workforce. QCAA staff themselves were relatively ambivalent on the extent to which QCAA processes could support the sustainability of the casual assessor roles.

Figure 9.48 QCAA staff confidence in sustainability of casual assessor workforce



In focus groups, QCAA staff reported that the major risk to the sustainability of the casual workforce was the workload of teachers.

> *But a lot of the things we get back from assessors is the workload. So they work for us. They work for their school Potentially they're doing after school camps and excursions and stuff like that, so it's really hard for us to determine our own workforce. When our workforce is relying on the schools, not overworking them first because we're second employment. So if they're going to have to drop the ball on something, it's probably going to be us. It's not going to be their primary job, so there is work in there to really try and tease out how do we keep these assessors engaged for the long term, especially when there's assessors who are that burnt out potentially by now and they were doing it on good.* (QCAA staff member)

This was further supported in interviews with other stakeholder groups. Although teachers were motivated to take on casual assessor roles, there were concerns about the sustainability of the current model, considering teachers' heavy workload.

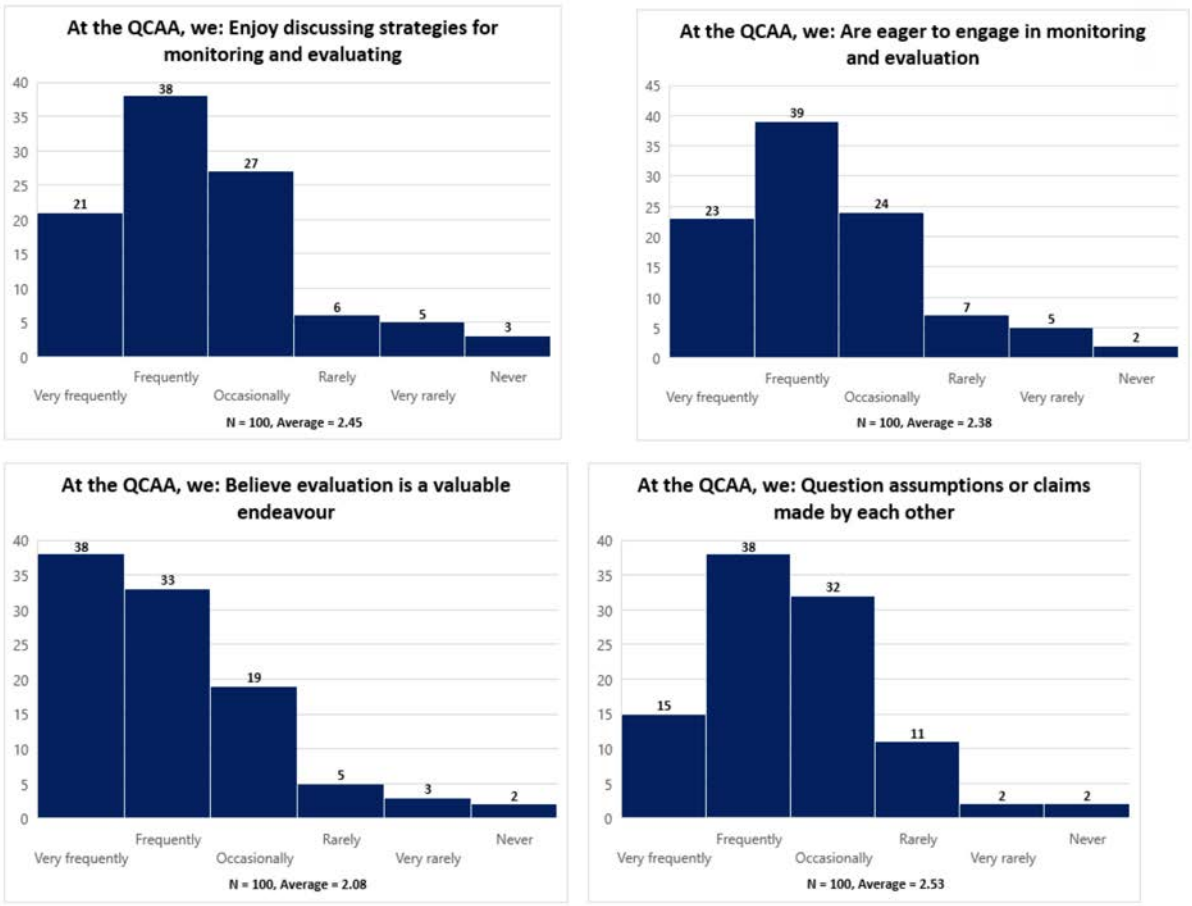
> *The involvement of teachers in the senior assessment process gives them influence and agency, and they're very committed to it. So, when we are saying things could be better, I know our members believe in teacher led assessment. So, they really do believe in it and they're flogging themselves. And the question that has to be asked is you know, how much do you want to flog them and how much do you want to recognise that it's part of their role and that they're contributing to their profession? They're not doing something else when they're over there doing that, they're doing their job.* (Teachers' Union representative)

> *The commitment to a quality assurance process and some level of moderation amongst the different schools again is another strong point of the QCE system. However, that particular process I think is in need of review, given the impact it has on work, workforce capacity and given everyone is experiencing staffing shortages. And whilst it provides fantastic PD for the individuals who perform the role of the Confirmers, Chief Confirmers, Chief Endorser … that places a lot of pressure on those schools that contribute to the good of the system overall by releasing staff to do so. I don't believe that's a sustainable model moving forward.* (Lutheran school sector representative)

The risk of unsustainability was considered even higher for subjects with few eligible assessors, such as languages or extension subjects. The lack of available teachers to fills these roles can also lead to other issues, such as conflicts of interest within the Confirmation process.

> *The pain point is that we don't have enough assessors to get through the workflow. So, I think the process of Confirmation, that part in the app, completing the review, applying the ISMG, providing feedback, etcetera is on point. It's that we just don't have the people power to maintain it... As languages, if there are two people away, sometimes that can be crippling to a subject, and it also questions, you know because the community of practice is so small. You've got lots of risks with leads, and chiefs may be accessing their schools or conflicts of interest, so they have to come out of the pool, and we have to look at another process, so they're not looking at their own schools.* (Confidential Interviewee)

Figure: 9.49 QCAA staff survey regarding discussing strategies, monitoring, and evaluation

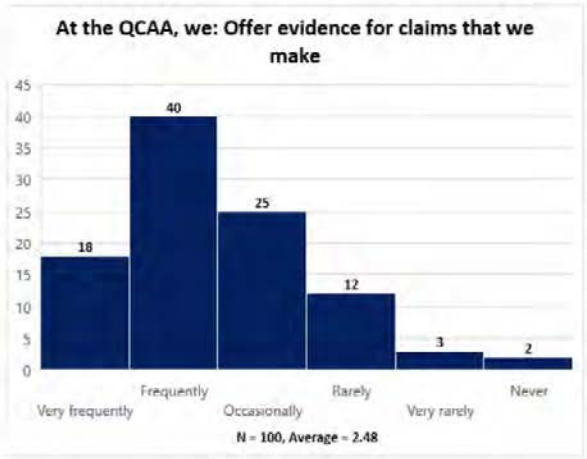


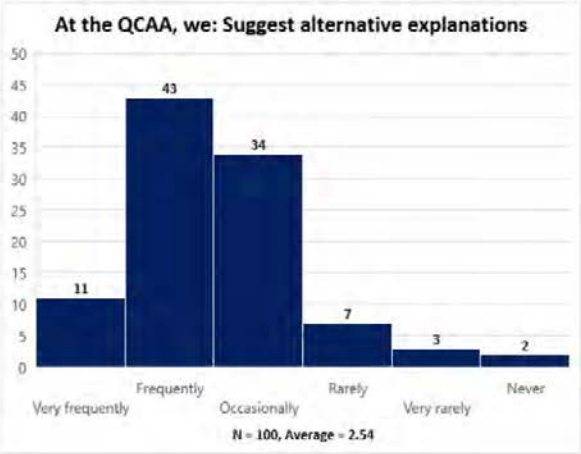
QCAA staff were further asked about their perceptions of evaluation, and the degree to which they believed it is a valuable endeavour. Again, most of the respondents (71 out of 100) reported positive attitudes to this, which may indicate staff see their organisation as a learning organisation where evaluation is considered a productive part of policy processes. A limitations of these survey responses is that respondents do not always do what they say or report they do (Choi & Pak, 2005); we therefore interpret these survey results as indications of an overall perceptions of the organisation and its policy.

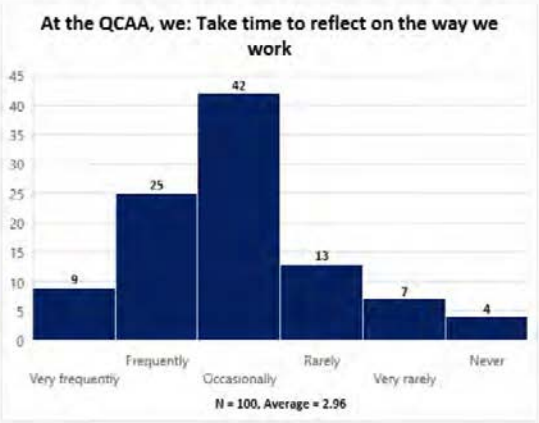
Interviews with key staff members in QCAA suggest that there is a very strong commitment to do the best for students in Queensland so that everyone can succeed. The dedication to change, and to being part of that change, was expressed across all categories interviewed. Some concerns were raised around the change process, given its complexity and growing number of staff in QCAA. It was indicated that even more documentation is needed on the change policy processes because of the growing number of staff in the organisation and the complexity of the organisation.

The survey responses indicate a culture in which most staff report they will offer evidence for claims they make, as well as being open to alternative explanations (see Figure 9.50 below). These responses somewhere more distributed and varied when respondents were asked whether they take time to reflect on the way they work, which could potentially be a reflection on the ongoing work pressure expected in an organisation like QCAA.

Figure 9.50 QCAA staff survey regarding evidence for claims, alternative explanations, and reflection







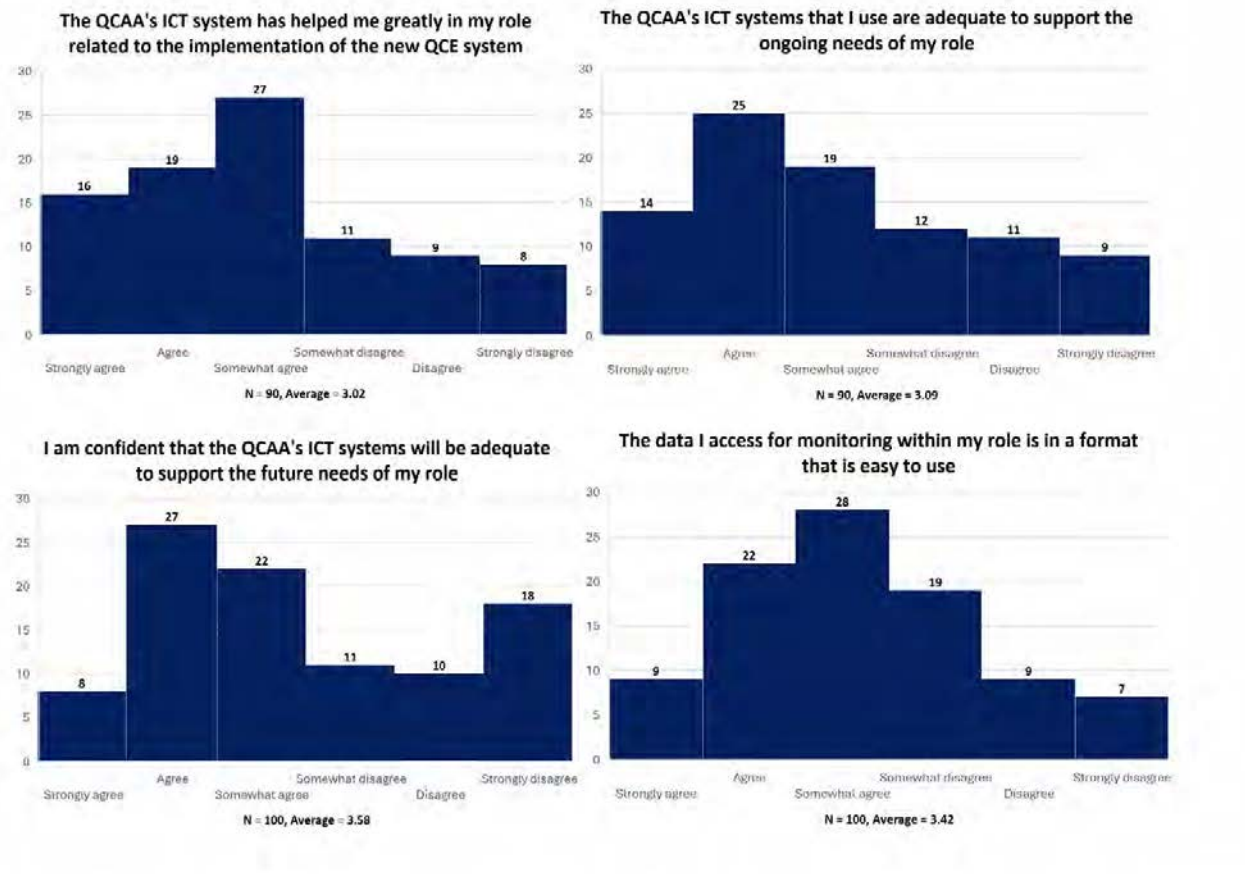
Interviews of external partners, such as the School Leader Forum, revealed appreciation of the ability to call QCAA staff and discuss policy matters; most external partners commented on how easy it was to speak to a person in QCAA who could help them solve problems. The external staff explained that the policy of being transparent around matters has increased trust. One example given was how QCAA chose a policy of transparency around and detailed explanation of an error in the 2023 Physics paper. The actions taken to resolve the issues and the associated resolution were openly discussed and explained, which built confidence in the organisation. Interviews with external partners also documented that there is trust in the policies in QCAA, with respondents described QCAA as an organisation that is responsive to changing context. The Junior Academic Integrity toolkit was used as an example of this. Further, QCAA updates and improvements to apps were appreciated. Data from the School Leader Forum pointed to factors outside of QCAA as the most pressing, such as teacher shortages and a perception that the Initial Teacher Education is not providing their Teacher Students the necessary assessment literacy. School Leader Forum staff suggested that more, and more targeted, professional development is needed for those teachers and principal delegates working in the QCE for the first time.

ICT Systems

The new QCAA portal, which is part of the new QCE system, serves as the central interface for ICT systems supporting the QCE framework. It is set up to be reliable and accessible for teachers and school leaders across the state, providing everyone with information needed independent of where you are teaching or go to school: '[R]egistered users log in to the QCAA Portal with a username and password that provides access to the applications, data and information for which they have permission. Each user is granted only the access that is relevant to their position in their organisation' (QCAA, 2025a).

As reported above, school leaders expressed appreciation for the new app and the fact that it is updated regularly. On the other hand, the most commented-on issue, at 16.7% of the comments provided, was around ICT systems – internal to the agency, facing the assessors and the QCAA portal. These comments are supported by the survey results, as the ICT items were negatively rated compared to most survey items across all groups. Staff survey results indicated some challenges in accessing data and the format available. Staff confidence in the adequacy of support for future needs was notably low, paralleling concerns raised by the Confirmation assessors.

Figure 9.51 QCAA staff views: ICT systems



Chapter 6 Findings: Monitoring, review, and system impact

Figure 9.52 QCAA staff views on monitoring and evaluation

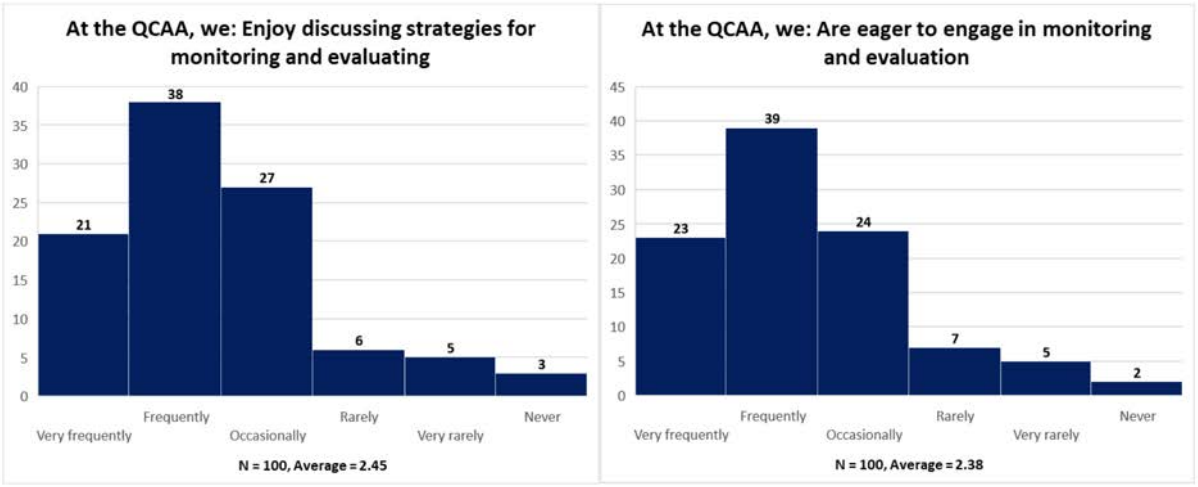


Figure 9.53 Principals' judgments on QCAA's responsiveness to school feedback

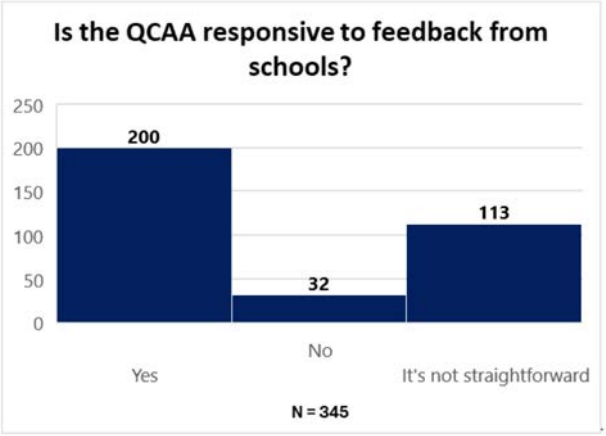


Figure 9.54 Teacher reports of data provided by QCAA

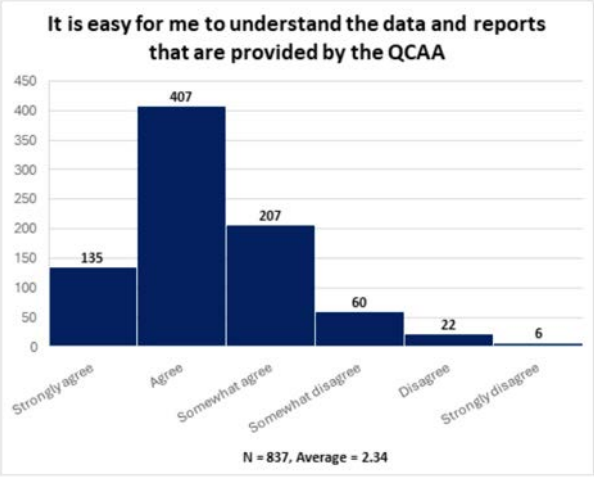


Figure 9.55 School leaders' attitude towards QCAA portal

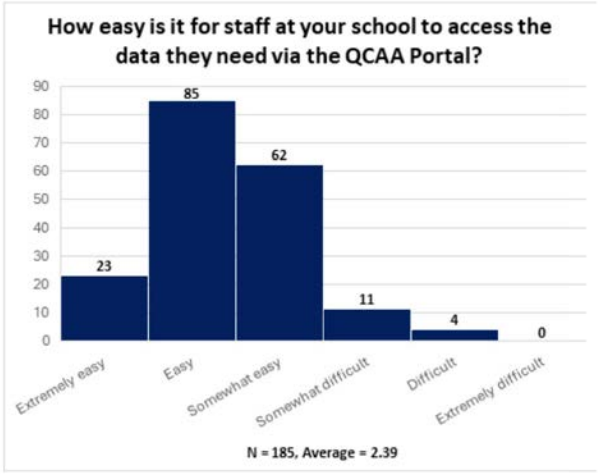


Figure 9.56 Teacher (left) and school leader (right) reports of data relevance

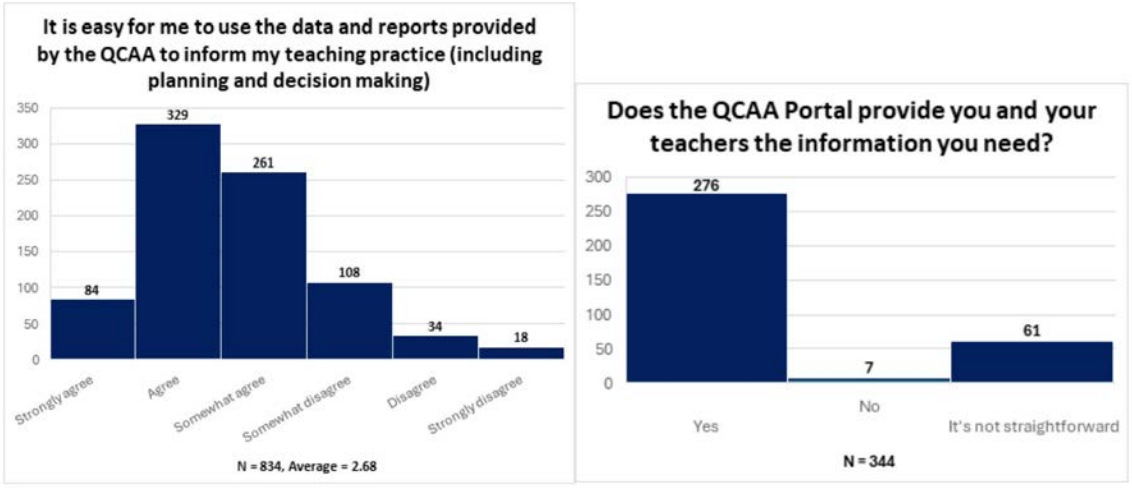


Figure 9.57 School leaders' assessments of the new QCE

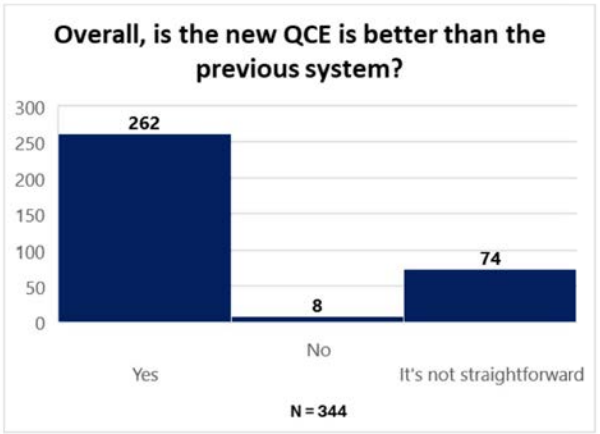


Figure 9.58 QCAA staff comments on the new QCE

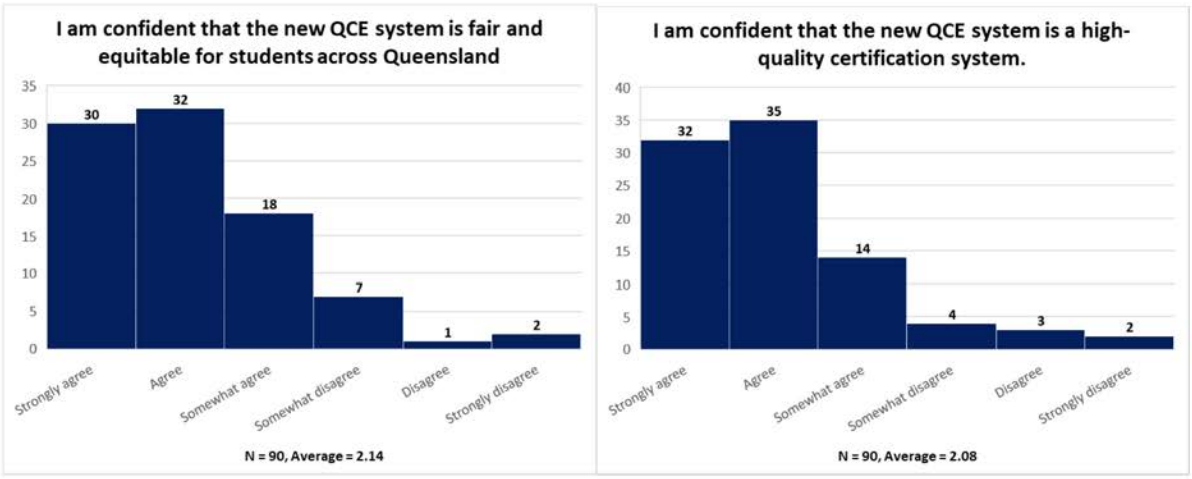


Figure 9.59 Invigilation and public confidence

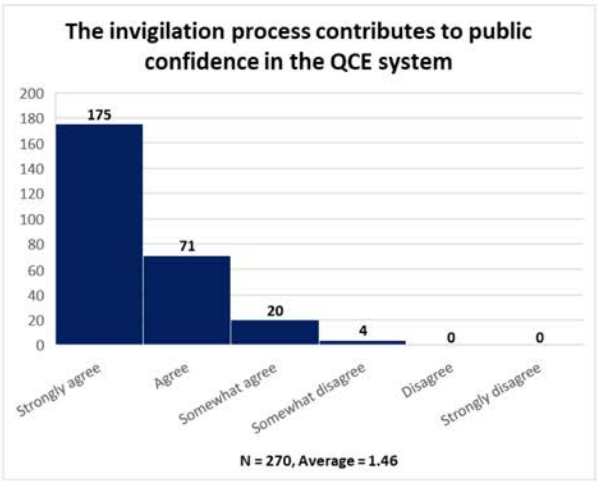


Figure 9.60 School staff comments on the new QCE

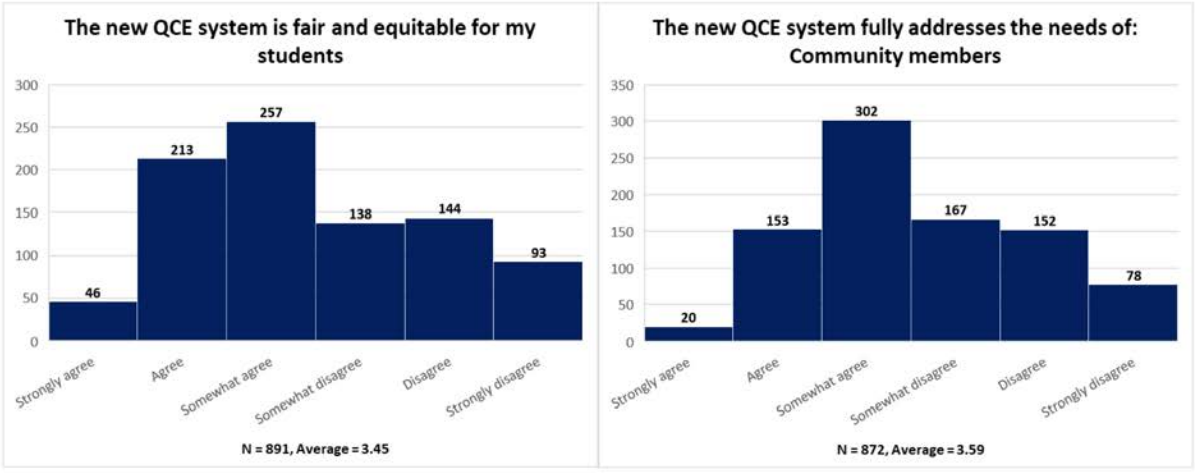


Figure 9.61 Invigilator perceptions of academic integrity during External Assessments

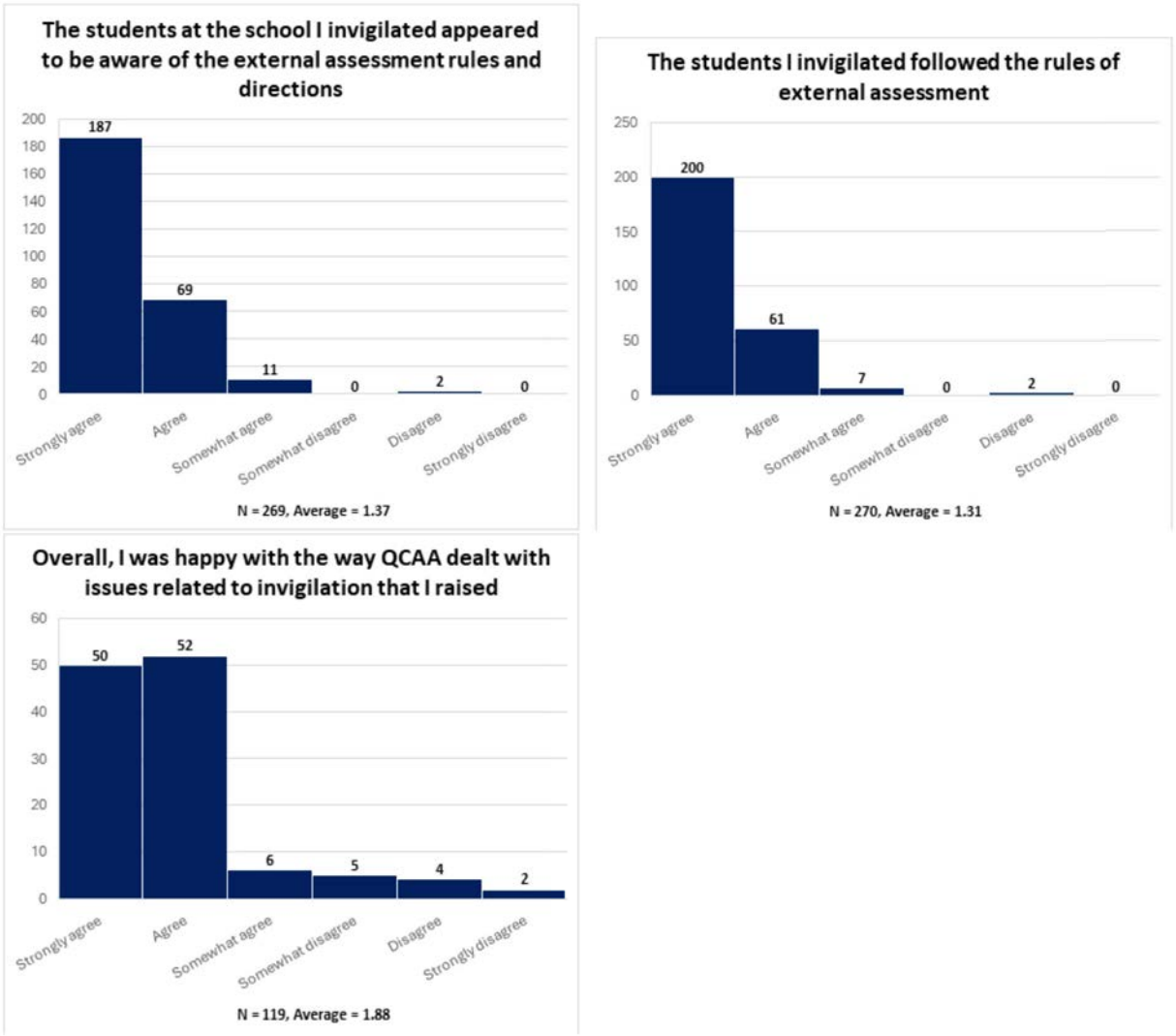
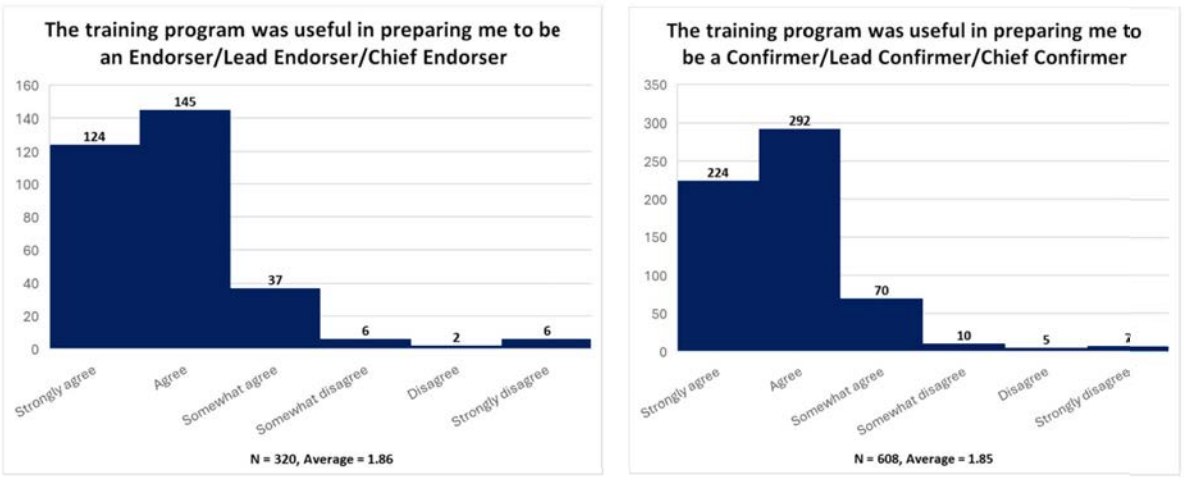
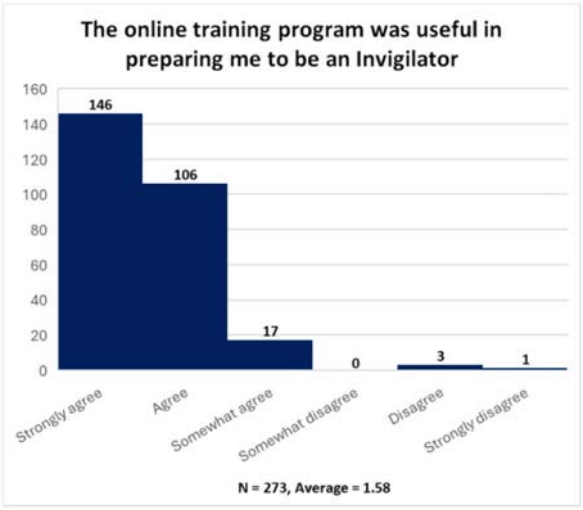


Figure 9.62 Attitudes of assessors and invigilators regarding the training program





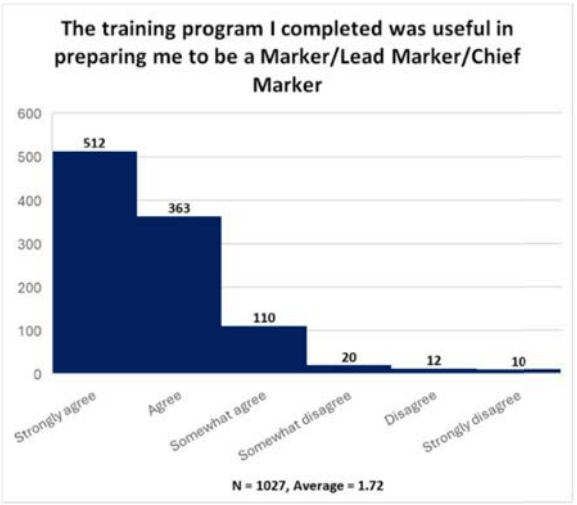
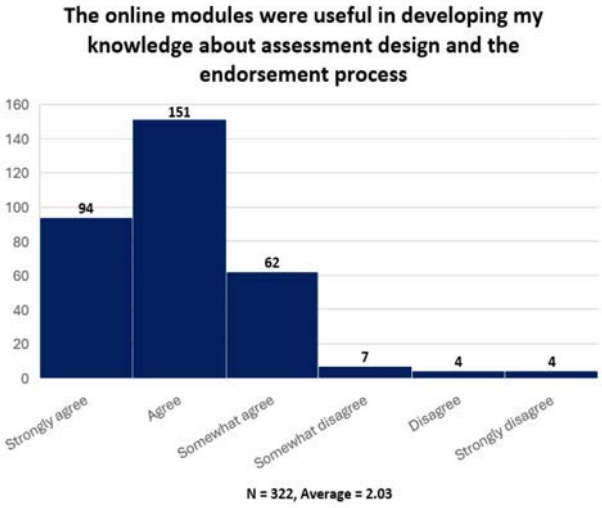
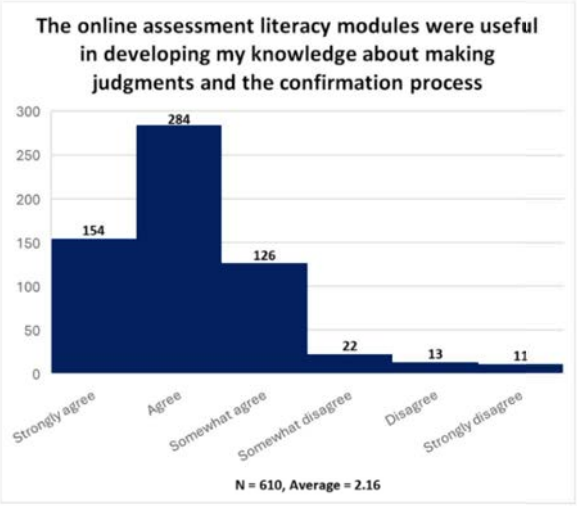


Figure 9.63 Attitudes of assessors regarding the online assessment literacy modules





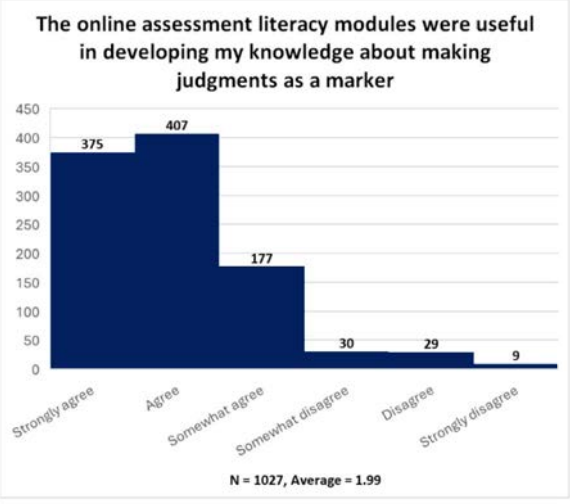


Figure 9.64 School staff members' satisfaction with professional development

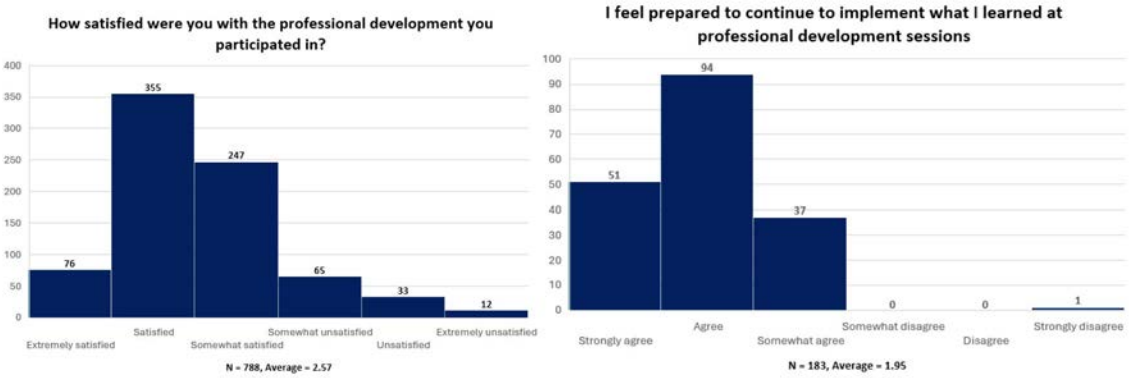


Figure 9.65 School leaders' attitude towards QCAA professional development and support

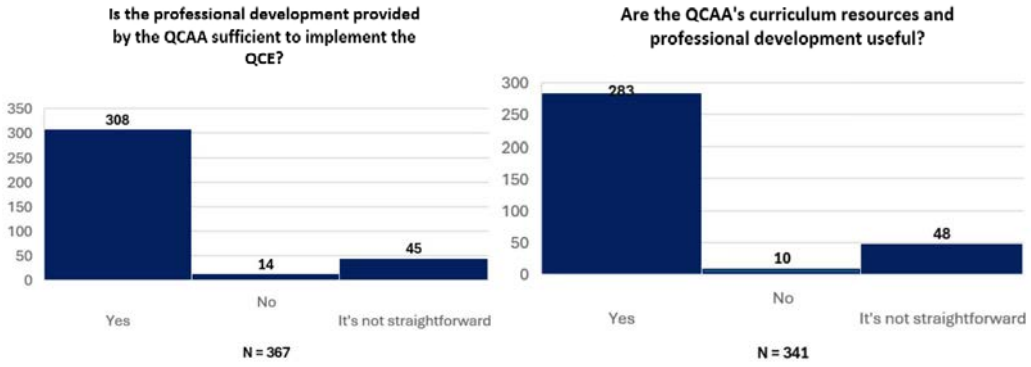
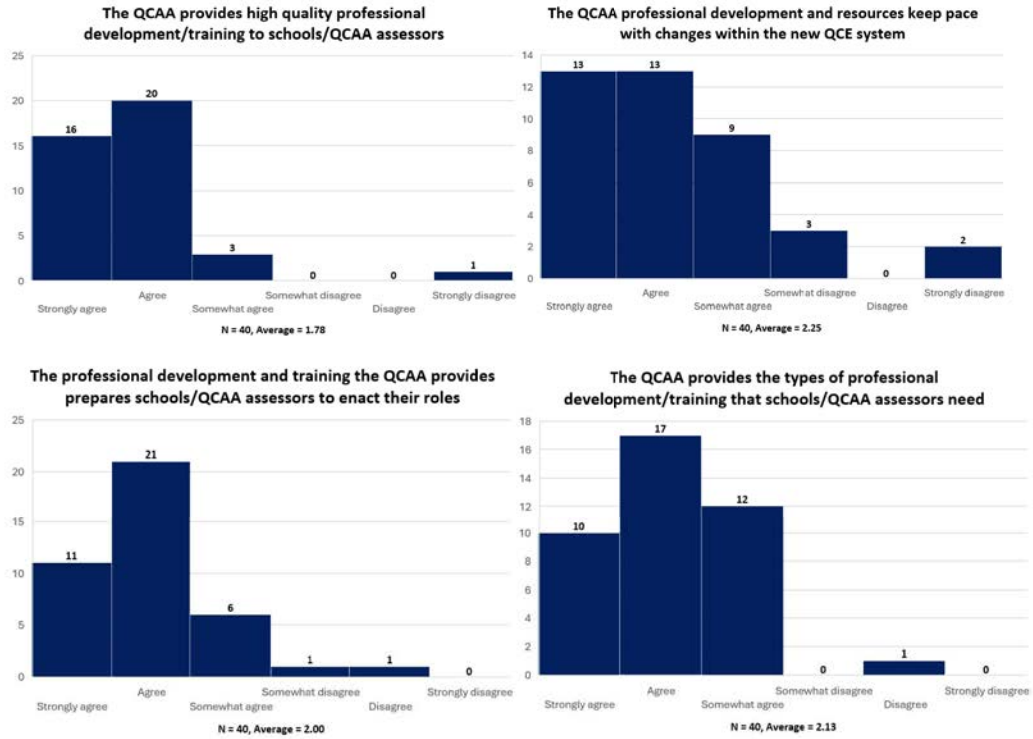


Figure 9.66 QCAA staff members' attitude towards QCAA professional development and support



THE UNIVERSITY OF
MELBOURNE