

YEAR-TO-YEAR COMPARABILITY OF RESULTS IN QUEENSLAND
STUDIES AUTHORITY SENIOR SECONDARY COURSES THAT INCLUDE
SCHOOL-BASED MODERATED ASSESSMENTS: AN ACACA SPONSORED
REVIEW

Scott F. Marion¹.
National Center for the Improvement of Educational Assessment

Bob Peck
Victorian Curriculum and Assessment Authority

Jan Raymond
SACE Board of South Australia

June 7, 2011

¹ Authors are listed in alphabetical order

Introduction and Charge

Comparability of results is always important when assessment, whether through public examinations or externally-moderated school-based assessment, is used for publicly defining and reporting students' achievements in the courses they study. In the context of senior secondary certification, results recorded on certificates should (unless specifically identified otherwise) be comparable from person to person, from year to year and from provider to provider. In this context, two results that appear on a certificate to represent the same achievement in the same course should in fact represent essentially similar achievement, even where the forms of assessment were different and the assessment took place in different years.

This report is part of a larger ACACA project exploring the feasibility, including costs, of various options for ensuring reasonable comparability, both year-to-year and across Australia, once ACARA has developed content and achievement standards for a range of senior secondary courses in English, mathematics, history and sciences such as physics and chemistry.

As one part of this ACACA project, this report is focused on a review team's evaluation of the processes used by Queensland to ensure an appropriate level of year-to-year comparability of results in externally assessed courses, particularly in terms of achievement levels across years for the same course. The review was designed to evaluate the processes and procedures used by Queensland to ensure within-course, across-year comparability. To conduct this review, the within-year comparability of results across schools in the state was considered as part of the overall quality control and quality assurance system.

Method

A three member review team was nominated by ACACA to conduct the reviews in each of three jurisdictions: Tasmania, Queensland, and New South Wales. Two senior experts from qualifications authorities not being reviewed (Victoria and South Australia) and one assessment and accountability expert from the United States served as the reviewers in all three jurisdictions (see Appendix A for a short biographical sketch of each reviewer). The composition of the review team was selected to provide a suitable combination of peer and external review.

The one-day onsite review included a presentation from the qualifications authority on the current comparability procedures with substantial time for the reviewers to raise questions based on the documentation and/or the issues discussed in the presentation. The first part (generally about one-half of the time) of each onsite review was devoted to familiarizing the reviewers with the state's senior assessment and certification system and gaining a solid understanding of the within-course, within-year, across school comparability. In addition to hearing presentations from senior officials in each jurisdiction, the reviewers relied on the prepared documents and other background knowledge to raise clarifying questions to make sure that they were clear about how the system worked.

The Queensland Studies Authority sent the following documents to the review team about two weeks in advance of the review:

- ✓ A-Z of Senior Moderation
- ✓ School-based assessment: The Queensland system (booklet)
- ✓ Externally moderated school-based assessment: Myths and facts
- ✓ Calculating OPs: The basic principles
- ✓ Queensland Core Skills (QCS) Test Handbook 2010
- ✓ Queensland Core Skills Test retrospectives and multiple choice response sheets
- ✓ Random sampling reports
- ✓ State Review panel reports
- ✓ Comparability processes for Authority subjects: Systemic consistency and reliability
- ✓ Queensland comparability study: 2009 and 2010

The following QSA personnel were interviewed as part of this review:

- Peter Luxton, Acting Director
- Peter Jordan, Acting Deputy Director, Assessment and Reporting Division
- Brian Nott, Assistant Director, Analysis and Reporting Branch
- Leanne Rolph, Acting Assistant Director, P-12 Implementation Branch
- Natalie Carrigan, Acting Manager, Quality Assurance Unit

Once the reviewers clearly understood the within-year system, the review focused on describing how the system worked across years to ensure that students taking the same course are held to the same standards across years. The reviewers examined the current processes and procedures used to ensure year-to-year comparability and importantly examined the results of any studies, published or internal, that evaluated the validity of the year-to-year scores. This included a self-assessment by the authority staff on the strengths and challenges of the current comparability evidence as well as plans to continue to evaluate the comparability of year-to-year results.

The review was structured to afford time towards the end of the day for the reviewers to meet alone in order to consolidate their reactions and frame the “report out” to senior QSA leadership. This time also provided the reviewers with the opportunity to ensure that they clearly understood the system and if questions remained, they were addressed in the final session with authority staff. The review finished with the review team presenting its initial findings to the Acting Director of the authority and key staff members. These findings were presented using the following organizational scheme:

- ✓ **Commendations:** Current QSA practices that the reviewers found to support within-year and year-to-year comparability inferences.
- ✓ **Affirmations:** New, beginning or planned practices to ensure/improve year-to-year comparability of the QSA that the reviewers wanted to encourage the authority to continue.
- ✓ **Recommendations:** Suggestions by the reviewers for the authority to consider in terms of processes for improving comparability or for studies to help evaluate current and future comparability inferences.

In several places, the reviewers had additional questions, but these questions generally related to an already stated commendation, affirmation, or recommendation. Therefore, questions are included in several of these statements below.

Findings

In general, the reviewers were favorably impressed with the efforts of the QSA to maintain the quality and year-to-year comparability of the senior secondary assessment results. The detailed findings about the system are presented according to the organizational scheme described above.

Commendations

1. **The QSA has produced an outstanding set of documentation describing and documenting the quality of the externally-moderated school-based assessment system.** The documentation ranged from user-friendly pamphlets to sophisticated statistical reports. QSA has a long program history and has had the opportunity to develop this documentation over the years. Yet, QSA clearly updates these documents regularly and has indicated an ongoing commitment to producing new documents on a regular basis that reflect up-to-date analyses and program policies.
2. **QSA, through its documentation and interview responses, has demonstrated clear evidence of a commitment to learning about and supporting their assessment system and the ACACA comparability project.** The set of recent analyses and documents which evaluate the comparability of the Queensland assessment results is further evidence of this ongoing commitment to continuous improvement.
3. **The training, support, and monitoring of panelists at all levels—school, region, and state—is quite strong.** Further, the multi-level structure of the panel reviews adds strength and credibility to the system. The Queensland system rests on the expertise and commitment of the panelists (who conduct moderation meetings) and it is clear that QSA recognizes and supports the need to maintain a high level of expertise among its panelists.
4. **One of the most important and impressive strengths of the Queensland system is that comparability is not established through some *a priori* comparison of assessment forms, but through an intense focus on actual evidence of standards demonstration from student work.** Assessment results do not depend on what is purported to be measured, but on what knowledge and skills are actually demonstrated through actual student evidence.
5. **As has been well-documented in several places, the externally-moderated, school-based assessment system is designed and implemented to both build capacity in and empower local educators.** In addition to the strong locus of control at the school level, the capacity-building and empowerment is enhanced through the strong feedback loop and processes inherent in the QSA approach.
6. **The syllabus and work program approval processes are very important to the success of QSA’s school-based assessment approach.** Given that the assessments are all situated within courses, it makes sense to check that the proposed approach for teaching the course meet QSA expectations. This avoids the potential harmful situation

where assessment results submitted for a course from a particular school are deemed unacceptable, in large part because the course itself was designed inappropriately. While checking the syllabus and work program ahead of time does not guarantee against this outcome, it is an important safeguard.

7. **QSA’s systematic approach to training all 54 subject panels is a strong feature of the system to ensure that there are common expectations across all subject panels.** This common training scheme will not necessarily lead to comparability of awards across subjects, which is not necessarily the intent, but it will help communicate generally common expectations for students and teachers within schools and should support the overall school-based assessment system.
8. **The “random sampling project” and analyses is a useful component for ensuring comparability.** The random sampling project is a post-hoc set of analyses designed to evaluate the degree to which the school-based, regional, and even state panels are adhering to and maintaining standards. The reviewers thought that the random sampling project was a useful addition to the state’s comparability processes and encourage QSA to continue to work on refining how the results of the random sampling analyses are fed back into the moderation processes to improve the year-to-year comparability.
9. The main approach for ensuring within-year and year-to-year comparability, as we have been discussing, is the externally-moderated system that relies on careful scrutiny of student work samples. As powerful as this approach is, **QSA is to be commended for supplementing year-to-year comparability with a variety of statistical analyses.** Some of the statistical analyses were used to determine if any adjustments were necessary before results were released for that year, but given the tight timeframe for producing yearly results, the main function of the analyses was to reflect on the processes in order to make adjustments in subsequent years. Trying to make adjustments within the same year on such a compressed schedule could lead to errors. More importantly, the QSA approach allows for a more deliberative reflection on the overall processes than if the Authority was trying to make “in the moment” adjustments in the current year. That is, the statistical analyses, particularly the random sampling project, were an important check on the external moderation results and used to support the continuous improvement process.

Affirmation

1. **The reviewers thought that QSA’s digital training system used to supplement the face-to-face training (which is only available every second year) was a great addition.** The reviewers recommend continuing to support the online digital training, especially for panelists joining in odd years. The digital training does not have to be used simply for odd years, but can be used to increase the reach of the training in all years and for capacity building within schools at any time.
2. **The web-based (especially) and paper materials developed by QSA for supporting assessment literacy are very impressive and appear promising in how they might support better student outcomes.** While the reviewers could not gain a full understanding of how these materials are used and understood by the field, it was clear to the review team that the materials were high quality and would likely promote effective classroom and school-based assessment practices.

3. **The comparability summary reports produced by the state panel chairs are a step in the right direction for providing feedback to schools and the education community generally.** The State Review Panel Reports are a useful summary of subject-specific information with respect to judgments of standards and comparability issues. In order to foster improvements in the district panel processes, the ACACA review team supported sharing these reports with the district panels, perhaps with additional interpretative information, to improve the quality of feedback on district moderation processes.
4. **In general, the review team was encouraged by the various forms of year-to-year comparability analyses such as the random sampling project, and support QSA's efforts in pursuing improvements in year-to-year comparability analyses as a way to feed into within-year and year-to-year comparability improvement.**
5. The report, *Queensland comparability study: 2009 and 2010*, presented a great deal of interesting information. Having the various award levels line up on the identity line indicates that the relative difficulty (standards) is maintained within each subject across years. The reviewers were interested, as well, by the apparent differences in "difficulty" across subjects, but recognize that comparisons across subjects are not the main function of these analyses. This report led to questions among the reviewers about the appropriate comparability criteria, but, by and large, the reviewers appreciated the intentions of these analyses and encourage QSA to continue to pursue such studies.

Recommendations/Questions

1. **The reviewers questioned whether it makes sense to do the Year 11 review every year?** In other words, the reviewers thought it might be more efficient, without a loss of effectiveness, by reviewing half the schools every year, for example? On a similar note, the reviewers wondered if it would help improve the efficiency of the system if schools could be "accredited" for some number of years based on consistent acceptable performance.
2. **While the reviewers thought that QSA engaged in many thoughtful approaches for evaluating the year-to-year comparability, the Queensland system could benefit from "seeding" work from previous years into the current year's moderation processes.** This would be a direct approach to evaluate the degree to which standards are maintained across years. The reviewers recognize that there might be some logistical challenges to overcome in order to carry out such studies, but such studies could prove so useful that the reviewers argue it is worth trying to overcome the potential hurdles. However, given the tight timeframe for end-of-year certification processes, the reviewers suggest that the proposed "seeding" studies are best carried out as part of the random sampling project completed in the first few months of the following year.
3. **The reviewers appreciated seeing draft reports that will be used to provide feedback to panel members, but suggest that QSA work to make the data collection, reporting and feedback on panelist performance more systematic across the districts and state.**
4. **While the random sampling reports are displayed fairly prominently on the QSA website, the reviewers suggested preparing reports that expand on the analyses presented in figures 6 and 7 in the 2010 random sampling report, for example, to provide clear and unambiguous feedback to district panels.**

Summary

This report describes the review process and findings of a small external review of QSA's senior secondary assessment process, particularly in terms of the within-subject, year-to-year comparability of assessment results. As noted above, the reviewers agreed that the current procedures for ensuring both the within-year and year-to-year comparability were strong overall. The quality of the specific personnel involved in the multiple levels of review panels is particularly strong. There is a significant institutional knowledge base residing in these personnel, which is a significant source of strength for the program. As noted above, the reviewers appreciated the intense efforts of the QSA and the school-based educators in Queensland to create a system with a primary goal of supporting high quality instruction while being able to support a critical accountability function. Queensland has created an internationally respected model of assessment and the policy leaders in Queensland should be proud of their success.

Appendix A: Review Biographies

Scott Marion, Ph.D. is the Vice President of the National Center for the Improvement in Educational Assessment, Inc., a U.S. based non-profit consulting firm. His current projects include developing approaches for states leaders to use in designing and implementing validity evaluations for large-scale assessment systems, exploring the instructional usefulness of interim assessments, and helping states design valid assessment and accountability systems. Dr. Marion coordinates and/or serves on seven state technical advisory committees and is an expert panelist for three major national initiatives. Dr. Marion is a regular advisor to the United States Department of Education (USED) on a host of assessment and accountability issues and is a member of USED's National Technical Advisory Committee. Dr. Marion was one of a limited number of assessment experts asked to provide extensive testimony and advice to USED regarding the Race-to-the-Top assessment competition. Dr. Marion has recently served on two National Research Council (NRC) committees investigating the issues and challenges associated with incorporating value-added measures in educational accountability systems and the second on outlining the "best practices" in state assessment systems. A former field biologist and high school science teacher, Dr. Marion earned a Bachelor's degree in biology from The State University of New York and a Master's in Science Education from the University of Maine. Dr. Marion received his Ph.D. in measurement and evaluation from the University of Colorado, Boulder. Prior to joining the Center for Assessment in early 2003, Dr. Marion was most recently the Director of Assessment and Accountability for the Wyoming Department of Education and was responsible for overseeing the Wyoming Comprehensive Assessment System and designing the technical and policy structures to implement the Body of Evidence assessment system, a multiple-measures, locally-created collection of evidence used to determine whether high school students met the state graduation requirements. Dr. Marion regularly presents the results of his work at several national conferences (AERA, NCME, and CCSSO) and has published dozens of articles in peer-reviewed journals and edited volumes.

Bob Peck is the Senior Technical Specialist at the Victorian Curriculum and Assessment Authority, a statutory authority that provides high quality curriculum, assessment and reporting for students in the state of Victoria, Australia. Having received a D.Phil. in chemistry from Oxford University he taught science for 17 years in secondary schools in Fiji, England, and Australia. After this, he spent three years at Murdoch University doing research on the zinc-bromine battery, concurrently acquiring tertiary teaching experience in education, chemistry and the social sciences. For 14 years he worked at the Curriculum Council of Western Australia (formerly the Secondary Education Authority) where he was eventually Senior Consultant, Measurement and Research. In this role he was responsible for carrying out examination development, test analysis and educational research. In 2000, he completed a doctorate in education at Murdoch University under the supervision of Professor David Andrich, one of the world's leading experts in psychometrics. He has been a consultant to Edith Cowan University, the Albanian Education Reform Program, the Fiji Education Sector Program and the South Pacific Board for Educational Assessment. In 2003 he moved to the Victorian Curriculum and Assessment Authority, initially as Manager of the VCE Examinations and then in 2009 to his current position as Senior Technical Specialist.

Jan Raymond is the Executive Manager, Moderation Services of the SACE Board of SA. Her background as an Assessment Field Officer for the SACE Board followed a significant teaching career as a specialist teacher of mathematics and geography, as a middle manager in schools, and as a moderator of mathematics in South Australia.

Jan's management experience was furthered through leadership positions in the Curriculum and Assessment Branch of SSABSA, as it was then known. More recently she led major structural change within the organisation as Executive Manager, HR and Corporate Services to meet the outcomes of the SACE review (2006) and major legislative changes under the SACE Act (2008) governing the organisation and its activities.

More recently, Jan took up the position of Executive Manager, Moderation Services in which she has been responsible for: the policies and overall management of the quality of school assessments for the SACE; managing quality assurance procedures for the consistent application of performance standards in Stage 1 and Stage 2 school-based assessments; and providing assistance to schools to strengthen school-based assessment practices including:

- providing feedback/information to schools at key points of the assessment/moderation cycle to strengthen the use of performance standards;
- using statistical systems to monitor and oversee the investigations into significant variations between the external and school-based assessment of Stage 2 subjects;
- managing the Assessor Training program.