

**Record of Assessment Moderation Practice (RAMP): Survey software as a mechanism of
continuous quality improvement**

Genevieve Marie Johnson

School of Education, Curtin University, Perth, Australia

GPO Box U1987, Perth, Western Australia, 6845. g.johnson@curtin.edu.au

Receiving a doctoral degree from the University of Alberta (Canada) in 1990, Genevieve Johnson has been actively involved in university research and teaching for more than 20 years. She is currently Associate Professor and Director of Teaching and Learning in the School of Education at Curtin University.

Record of Assessment Moderation Practice (RAMP): Survey software as a mechanism of continuous quality improvement

In higher education, assessment integrity is pivotal to student learning and satisfaction and, therefore, a particularly important target of continuous quality improvement. This paper reports preliminary development and application of a process of recording and analyzing current assessment moderation practices with the aim of identifying areas in need of improvement. Specifically, survey software was used to create a record keeping strategy in which unit coordinators documented the assessment moderation and integrity practices in each unit during each study period. Such an online survey approach to record keeping was amenable to data analysis with statistical software which facilitated identification of trends and anomalies. Instructional staff responded well to the initiative. As is typically the case with monitoring of behaviour, improvements in assessment moderation practices were immediately apparent.

Keywords: assessment moderation; assessment integrity; higher education; university students; continuous quality improvement

Introduction: Continuous quality improvement and assessment moderation

Continuous quality improvement is “a planned approach to transform organizations by evaluating and improving systems to achieve better outcomes” (Hunter et al. 2014, 1). In the 1920s, continuous quality improvement originated in the manufacturing industry to improve quality and productivity (Colton 2000). In the 1990s, the concept was increasingly applied to improve the quality of a range of human services organizations (Gamble 2011). Continuous quality improvement involves the systematic assessment of program implementation and short-term outcomes in order to improve service delivery and long-term outcomes. It differs from traditional program evaluation approaches in that it involves an iterative cycle of monitoring performance, identifying problems and potential solutions, and implementing changes, as well as the involvement of frontline and other staff in the improvement process (Dew and McGowan Nearing 2004). Although a variety of theoretical paradigms of continuous quality improvement exist, Rubenstein and colleagues (2014) identified three essential features: 1) systematic data guided activities, 2) iterative development and testing and 3) designing with local conditions in mind.

Continuous quality improvement in higher education is based upon the assumption that processes and products can and should be improved over time (Dew and McGowan Nearing, 2004). New tools provide new opportunities for improving educational and administrative processes and increasing outputs, that is, students can always learn more (Gamble 2011). In higher education, assessment exerts a profound effect on student learning and satisfaction (Klenowski 2011). Indeed, university grades are of considerable consequence in terms of subsequent student employment and educational opportunities. Since the importance of valid student grades cannot be overstated, assessment moderation and integrity processes constitute particularly critical targets of continuous quality improvement in higher education.

Assessment moderation in higher education refers to “a process for assuring that an assessment outcome is valid, fair and reliable and that marking criteria have been applied consistently” (Bloxham 2009, 212) and as such is a critical factor in assessment integrity (i.e., the validity of assessment marks as indication of student learning). Assessment moderation is described as “a practice of engagement in which teaching team members develop a shared understanding of assessment requirements, standards and the evidence that demonstrates differing qualities of performance” (Adie, Lloyd and Beutel 2013, 1). Wallace et al. (2010) reported that assessment moderation is generally conceptualized as involving accuracy, consistency, and comparability of marking, although precise definitions vary across instructional situations (Adie, Klenowski and Wyatt-Smith 2012). University assessment moderation strategies minimize the effect of individual assessor bias, satisfy government accountability demands, increase trust in standards, and improve the quality of student assessment tasks (Annetts, Jones and Deursen 2013; Bloxham 2012; Klenowski 2011; Smith 2012).

Although labour-intensive and associated with delay in student feedback (Buglear 2011; Nuttall 2007), assessment moderation has been associated with a range of positive learning outcomes (Adie et al. 2013). According to Bird and Yucel (2013), assessment moderation allows lecturers to provide reliable and consistent feedback that contributes to student learning. Smith (2012) argued that assessment moderation practices motivate students to learn because measures are put in place to ensure the fairness, validity and reliability of assessment marking. Similarly, Sadler (2009) maintained that assessment moderation has “a substantial affective impact on learners and their learning, influencing both students’ sense of achievement, and their motivation and level of engagement in future courses” (159). Klenowski and Wyatt-Smith (2010a)

contended that assessment moderation informs teaching, facilitates students' learning and promotes achievement standards in education and, therefore, contributes to "quality teaching and learning experiences" (110.) Further, assessment moderation reinforces standards which directs student focus on relevant learning content (Bloxham, Boyd and Orr 2011; Klenowski and Wyatt-Smith 2010b).

University assessment moderation practices

Assessment moderation is often equated with double marking, that is, two markers grade the same assessment (Smith 2012). While double marking is a common form of assessment moderation and appears to increase inter-marker reliability (Kuzich, Groves, O'Hare and Pelliccione, 2010), it is lamented for presenting a "singular or narrow view" of moderation (Adie et al., 2013, 2). In fact, double marking as a form of moderation is criticised for lacking reliability, as two assessors often have different reasons for giving a student the same mark (Sadler 2013). Annetts et al. (2013) argued that the correlation between the marks awarded by two assessors tends to be low. The total mark for an assessment may be similar across the two markers despite differences in marks awarded to assessment sub-sections, which gives double marking a false sense of high reliability. Added to this, double marking is said to be problematic because it fails to take into consideration factors such as assessor's various levels of expertise, marking fatigue, and marking time frame (Bloxham 2009).

In contrast to double marking which occurs at a single point in time, assessment moderation should occur across a number of assessment stages, -- that is, from planning and operationalization of assessment design to review of students' results (Mahmud et al. 2010). According to Kuzich et al. (2010), the assessment moderation process consists of six steps: 1) design of the assessment by the unit creator; 2) development of marking criteria and distribution to staff and students; 3) students' submission of assignments; 4) assessors' grading of assignments based on marking schemes; 5) assessors' discussion and validation of marks; 6) and return of assignments to students. Similarly although with more parsimony, Thuraisingam et al. (2010) proposed three phases of assessment moderation: 1) assessment design and development; 2) assessment implementation and marking; and 3) review and evaluation of marks. Moderation, in this "whole-of-course approach" (Smith 2012, 48), includes initial conceptualisation of the assessment tasks and active involvement of all assessors who discuss, negotiate and clarify

understanding of the marking criteria. This viewpoint is in accordance with that of many of the lecturers in the Wallace et al. (2010) study, who suggested that moderation must ensure that “assessment standards as well as content and delivery methods are well understood by all staff and students from the very beginning” (4). In fact, the Australian Tertiary Education Quality and Standard Agency has ruled that assessment moderation should be part of curriculum design, not just concerned with performative outcomes (Adie et al. 2013).

Related to the holistic view of assessment moderation is a widely supported strategy referred to as the *analytical criterion-referenced approach* in which each critical aspect of the assessment is judged independently as is commonly the case with the use of marking rubrics (Kuzich et al. 2010; Goos and Hughes 2010; Hunter and Docherty 2011). Wyatt-Smith, Klenowski and Gunn (2010) maintained that the strength of this approach is that it provides “explicit provision” (61) to support assessors’ marking decisions. Sadler (2013) asserts that “judgements should be made in as direct and absolute a way as possible” (14). Bloxham (2009) and Orr (2010) point out several benefits of the criterion-referenced approach to moderation, which include improved reliability and objectivity, avoidance of the effects of soft markers and hard markers, and development of staff confidence in marking assessments. In a study comparing assessment with criteria and without criteria, Van der Schaaf, Baartmanb and Prins (2012) found that, when marking without criteria, assessors relied mainly on their personal opinions and less on students’ quality of work, whereas when marking with criteria, their judgement processes were more informed and the quality of their marking was enhanced. Bloxham et al. (2011) pointed out that “using criteria is considered best practice to the point that they are mandatory in some universities” (656).

However, criterion-referenced approaches to assessment moderation are not unanimously endorsed. In a study of 12 assessors in two universities in the UK, Bloxham et al. (2011) found that most assessors made holistic rather than analytical judgements without “evidence of linear or discrete processing of individual criteria” (662). Furthermore, a majority of assessors revealed they referred to the marking rubrics as a post-hoc process to justify their holistic judgement. Instead of using the provided marking criteria, assessors developed their own “standard frameworks” based on their knowledge of students’ work (655). Added to this, some of the assessors tended to work backwards, making a holistic decision of the total score first, and then assigning marks for each criterion (Bloxham, 2012). Annetts et al. (2013), in exploring the

marking of 75 dissertations, found that some examiners “relied mainly on experience” (314), and did not base their judgements on the provided marking rubrics. Similarly, Sadler (2013) found that less proficient assessors tended to rely more on marking criteria while more experienced assessors tended to make holistic judgements. Adie (2012) reported that some assessors weighted the criteria, combining “marks in different ways which were not part of the policy discourse” (51). Bloxham (2009) observed discrepancies between the marking results and the marking criteria. Hunter and Docherty (2011) also found that assessors’ comments about students’ performance against the marking criteria were at odds with the overall grades. Sceptical about assessors’ use of criteria, Wyatt-Smith et al. (2010) maintained that, despite explicitly stated marking standards, it is unclear how “teachers ascribe value and award a grade to student work” (59). Ashworth, Bloxham and Pearce (2010) contended that assessment is a subjective activity, since assessors tend to choose their own interpretations of the marking rubrics based on their experience and “their differing tacit knowledge of disciplinary standards” (222). Ironically, while students are expected to observe the assessment criteria as closely as possible (Smith 2012), assessors do not seem to share a common understanding of criteria and standards.

Therefore, for the criterion-referenced approach to assessment moderation to be effective, it has been suggested that marking criteria should be socially constructed among assessors rather than predetermined (Smith 2012; Johnson 2013). In other words, as mentioned previously, assessors should discuss and debate to build shared-understanding marking criteria from the early stage of the assessment design. By doing so, assessment criteria can be modified to improve clarity and precision (Hunter and Docherty 2011). Moreover, course content, materials, resources, assessment tasks, model answers and marking guidelines have to be transparent to lecturers and students from the commencement of the course (Wallace et al. 2010). Bloxham et al. (2011) suggested that, not only lecturers, but also students should also be allowed to discuss standards and assessment criteria to improve the validity and reliability of assessment moderation.

Other strategies of assessment moderation include anonymous marking, social moderation, statistical moderation, cross-marking and expert moderation. Anonymous marking is a method of assessment moderation that helps to avoid the effect of the assessor’s knowledge of the students on the grades (Bloxham et al. 2011). In a study that examined the transcript of moderation conferencing sessions of 15 lecturers, Wyatt-Smith et al. (2010) found that many assessors

revealed their grading of students' work were reliant on their "observations of student attributes and dispositions" (71). Anonymous marking, therefore, may reduce the effects of such prior knowledge of students. As Orr and Bloxham (2013) stated, assessment "is not all about the student, instead it is all about the work" (248). Social moderation, also called consensus marking, involves groups of assessors meeting to discuss and review the consistency of their marking judgements to reach an agreement on the quality of the student's work, to avoid an impressionistic approach to marking and de-privatise teachers' marking practice (Sadler 2013). Klenowski and Wyatt-Smith (2010a) argued that all types of moderation where lecturers meet inevitably involve an element of socialisation, since assessors "interact with one another as they make explicit their judgements of student work samples" (116). Social moderation meetings are particularly important for new examiners to develop shared understanding of expected standards and therefore improve inter-marker reliability (Hunter and Docherty 2011). Statistical moderation, also referred to as scaling or linking, involves statistical calculation of students' results and comparing those to "an external anchor test or examination" (Maxwell 2010, 459). Cross-marking characterises assessors' remarking each other's marked assignments (Wallace et al. 2010). In high-stake situations where strong quality control is needed, expert moderation or panel moderation is required (Adie et al. 2012).

Research Questions and Case Study Approach: Toward a model of continuous quality improvement in university assessment moderation practices

Current survey software, commonly available in university contexts and familiar to instructional faculty, provides a simple mechanism of assessment moderation and integrity recording keeping in the complex and diverse contemporary university instructional environment. Correspondingly, data collected with survey software is easily analyzed with statistical software. Can survey software be used to collect information on assessment moderation practices? Can survey data be analyzed to determine areas in need of improvement in assessment moderation practices? Are university instructors amenable to use of survey software to monitor and improve assessment moderation practices?

Case study research consists of a detailed description of phenomena in the context of practice (Singh 2014). "The case study is particularly suited to research questions that require detailed understanding of social and organizational processes because of the rich data collected

in context” (Hartley 2005, 323). Thus a case study is not a method but, rather, a research strategy. The context is an intentionally aspect of research design. In organizational quality improvement, there are always be too many variable to control and thus experimental research is inappropriate.

This paper reports a case study of an application of survey software as a mechanism of continuous quality improvement in higher education. The term *survey* was avoided because it failed to capture the essential function of continuous quality improvement. Thus, the phrase *record of assessment moderation practice* (RAMP) was used with university faculty and administrators in all discussion and presentation of associated record keeping processes and products directed toward continuous quality improvement.

Record of Assessment Moderation Practice (RAMP)

The RAMP was trialed in a large university in Western Australia. Given the preliminary nature of implementation, only School of Education programs were included. In the context of case study university programs, a full-time student enrolls in four units per study period. Many units have multiple lecturers but all units have a single unit coordinator who is responsible for assessment moderation and integrity within the unit. In the case study programs, units were delivered on-campus, fully-online and/or internationally. In such an instructional context, the assessment moderation and integrity challenges were formidable albeit increasingly typical of the complexity of contemporary university course offerings.

In the context of the case study university programs, the RAMP consisted of unit coordinator response to a series of online survey software (Qualtrics 2014) items relevant to student assessment. A distinction between possible assessment moderation strategies for various instructional situations required slight tailoring of items and response-options, for example, in some cases fully-online unit coordinators were geographically distributed which rendered some moderation strategies unreasonable relative to on-campus unit coordinators with easy access to their instructional team. Thus, for the case study university programs, similar but not identical surveys were generated to annually collect data either twice (i.e., during traditional on-campus semesters) or four times (i.e., during 13 week fully-online study periods). However, consistency of items and response-options was generally maintained to allow for comparison of assessment moderation practices across instructional situations.

As the quality agenda in higher education has been at the forefront of policy in the case study university, unit coordinators are expected to engage in comprehensive assessment moderation practices. For approximately two years prior to RAMP implementation, instructional development workshops and teaching and learning forums on assessment moderation practices have been promoted to all instructional faculty, although not all unit coordinators attend such professional development opportunities. Such workshops and forums served to increase instructor awareness of assessment policy requirements and models of assessment moderation. Thus, implementation and associated assessment moderation terminology were consistent with organizational climate and required instructional expectations.

All semester two unit coordinators in the traditional on-campus case study university programs (n= 67) were sent an email requesting that they complete the required RAMP for their particular unit. Since unit coordinators were required to identify their unit from a list of all units offered during the semester, noncompliant unit coordinators were easily identified. Two follow-up email reminders were sent to unit coordinators who did not complete the required RAMP in the allocated one week: the first email reminder was sent at the beginning of the second week and the second email reminder was sent at the beginning of the third week. Sixty-three of the 67 unit coordinators completed the required RAMP. The email request to complete the RAMP as well as the final screen in the online survey requested unit coordinators forward to an administrative assistant required materials including “information provided to students that describes the assessments, marking rubrics or equivalent for each written assessment including tests, samples of marked assessments with completed rubrics (e.g., fail, credit, distinction, high distinction).”

The first page of the survey queried: 1) the unit coordinator university identification number which facilitated sending email reminders to complete the RAMP; 2) the unit being reported which was selected from a complete list of units offered in the case study programs during the semester; and 3) the number of summative student assessments in the unit which was necessary as subsequent items queried the details of each assessment as illustrated in Figure 1, and branched into other items depending upon the response-option selected. Some student assessment types did not require moderation due to objective marking as is the case, for example, with multiple-choice tests. Although formative assessment occurred, university policy does not require, nor is it common practice, to moderate formative assessments because marks do not

contribute to final grades. Some items went beyond assessment moderation process to include assessment integrity more generally, for example, plagiarism prevention strategies. As illustrated in Figure 2, dependent upon the response-option selected by the unit coordinator, textboxes appeared for purposes of explanation. Such textbox explanations allowed for provision of unit coordinators subjective rationale for not engaging in assessment processes assumed to decrease plagiarism and increase the reliability and validity of student marks via assessment moderation processes. Such information is useful to understanding the complexity of individual situations as well as illuminating areas of required improvement (e.g., increased plagiarism prevent strategies and assessment moderation processes).

-- Insert Figure 1 about here --

-- Insert Figure 2 about here --

Subjective marking associated with student written assessments and presentations were further queried with respect to specific approaches to assessment moderation. Based upon review of the literature and recognizing diversity in assessment moderation practices, a range of survey items were developed in order to capture all possible moderation strategies. Illustrated in Figure 3, unit coordinators reported current assessment moderation practice by *selecting all that apply*. The emphasis was on record keeping as a mechanism of continuous improvement as opposed to monitoring compliance with university policy, although the latter was also a consequence of RAMP completion by unit coordinators. The expectation was that all unit coordinators would complete a RAMP each semester or study period. Unit coordinators feedback on specific survey items and the processes of information collection was encouraged and forward to the Director of Teaching and Learning for the case study university programs. Such unit coordinator feedback ensured continuous improvement in processes and associated data validity and utility, and was promoted and encouraged based on such theoretical and practical assumptions. Consequently, data and subsequent analysis may vary slightly across study periods/semesters and perfectly aligned comparisons over time may not be possible in the case of some specific items that may change across context and/or over time. Thus is the complexity of continuously improving mechanisms of continuous improvement.

-- Insert Figure 3 about here --

An obvious advantage of survey software for RAMP completion was the ease with which collected data could be downloaded, analyzed with statistical software (i.e., SPSS for Windows),

and presented with graphics software (i.e., Adobe Illustrator). For example, with respect to the 63 completed surveys for the second semester units, one unit (1.6%) had one assessment, 21 units (33.3%) had two assessments, 28 units (44.4%) had three assessments, and 13 units (20.6%) had four assessments. Subsequent description of the assessment allowed for quick clarification of the apparent rogue unit with one assessment; it was a dissertation proposal unit. Additionally, Figure 4 provides further description of the case study first assessment which was of considerable interest to the university instructors and administrators for the obvious and perhaps extreme use of written submissions. Such assessment patterns were easily discernable with RAMP data and allowed for subsequent discussion with unit and program coordinators to encourage diversity in assessment strategy.

-- Insert Figure 4 about here --

With respect to the 63 completed surveys for the second semester units, unit coordinators further indicated that assessment moderation occurred in 51 cases and did not occur in eight cases (data was missing for four units which were easily identified for subsequent follow-up). The most common reason given for lack of moderation was that it was not required in multiple-choice tests and quizzes, although two unit coordinators maintained that moderation was not required due to the nature of the assessment (e.g., dissertation proposal or personal planning document) and two claimed that they cannot find a suitable co-moderator. Such reasons provide specific opportunities for improvement in assessment moderation processes by follow-up with the unit coordinators who made such comments in their RAMP. The focus of such discussion was on problem solving for continuous improvement in assessment integrity.

Figure 5 provides a visual summary of the nature and extent of moderation (i.e., *select all that apply*) for the 51 unit coordinators responding in the affirmative to the item *Is this assessment moderated?* Reported as a percentage of those who responded, one unit coordinator indicated that the first assessment was not aligned with unit learning outcomes. Such information allowed for subsequent critical review of the assessment and corresponding modification, albeit slight. Less than 40% of unit coordinators held a pre-marking meeting with the instructional team, although the reasons for this were not queried. Pre-marking discussion of rubrics is supported as a best practice in assessment moderation (Orr 2010; Sadler 2013). It may be that the instructional team had previously established familiarity with the marking criteria (i.e., from previous semesters). Less than 50% of unit coordinators reported randomly checking some

assessments marked by members of the instructional team; subsequent discussion revealed that unit coordinators felt that they were not allocated sufficient time in their workload to engage in all recommended assessment moderation practices. The workload issue, however, was not in relation to completing the RAMP but, rather, the actual time required to fully-engage in comprehensive assessment moderation practices. Nonetheless, such instructor feedback was useful in determining realistic expectations. Consequently, instructor perception of the workload associated with comprehensive assessment moderation was communicated to university administration.

-- Insert Figure 5 about here --

Further to Figure 5, Figure 6 depicts secondary analysis of the 51 unit coordinators who reported the nature and extent of moderation (i.e., *select all that apply*). The number of assessment moderation strategies reported was summed, a simple command in SPSS. As was readily apparent, considerable variability existed across units in the case study university programs. Three of the 51 units or 5.9% used two of the nine assessment moderation/integrity strategies; ten of the 51 units or 19.6% used all nine assessment moderation/integrity strategies. Such simple analysis allowed for meaningful discussion among university lecturers and administrators directed toward continuous quality improvement. Are all assessment moderation strategies weighted equally and thus appropriately summed? What are the unit coordinator and unit characteristics associated with various levels of assessment moderation strategy use? Do patterns of specific and collective strategies impact on student perceptions of assessment integrity and motivation to learn? RAMP data generated with survey software and analyzed with statistical software facilitated processes of continuous quality improvement which included benchmarking, identifying patterns and anomalies, seeking explanations and providing mechanism of positive change.

-- Insert Figure 6 about here --

Continuous Improvement of Record of Assessment Moderation Practice (RAMP)

Focus group discussion with program coordinators who manage unit coordinators revealed that preliminary implementation of the RAMP strategy was generally well-received by unit coordinators. The survey software allowed for examination and analysis of the time spend by each unit coordinator in completing the survey but such information reflects the time from start

until submission which, in some cases, was several days. This is, nonetheless, a positive feature of survey software since unit coordinators were able to complete the RAMP when convenient. In most cases, however, it appeared that RAMPs were completed in less than 20 minutes. It is likely that completion time would be further reduced as unit coordinators gain familiarity with the online data collection process. However, three of the 63 case study unit coordinators who completed the RAMP indicated that increased demands on their time must be recognized in their workload. Since completion was focused on continuous quality improvement, such comments were received positively by case study university programs administration. A cost-benefit analysis is required.

A lesson learned in preliminary application of the RAMP process was the importance of the accuracy of the list of units offered (i.e., first page of the online survey) and that the email requests to complete the RAMP be sent to all and only the appropriate unit coordinators. Analysis of data downloaded into SPSS for Windows required some initial manipulation due to errors of inclusion and omission. Apparently, some records of available units and associated unit coordinators were not meaningful in the context of RAMP completion for a specific semester (e.g., independent studies and dissertation units). Some unit coordinators who taught across programs (e.g., on-campus and online) included inappropriate units (i.e., on-campus and online units required a slightly different format) by using the *other* response-option from the list of units offered. In some cases, there were informal arrangements between instructional staff who both assumed they were unit coordinators for the same unit offered in different formats (e.g., on-campus and online). Such unit coordinator confusion and consequential need for some entries to be deleted prior to analysis could have been avoided by more careful consideration of: 1) the list of units on the first page of the interface and 2) the individuals who were sent an email requesting RAMP completion, -- extremely obvious in retrospect.

Case study application of the RAMP process resulted in numerous other such improvements to the actual online survey. For example, the first page of the online interface queried unit coordinators staff identification number which was subsequently used to determine who had not completed the RAMP and, thus, which unit coordinators required a follow-up email. Staff identification numbers, however, required conversion to staff name in order to determine staff email. A much improved approach in terms of minimizing steps in processes would be to have the unit coordinator select his/her name from a list of names, -- the same accurate list of

unit coordinator names used to ensure the initial request to complete the RAMP was emailed only to the appropriate unit coordinators. Additionally, although several items queried the source and type of co-assessors (not presented due to issues of case study program anonymity), no information was collected on the actual number of instructional staff involved in moderation processes for each assessment. Analysis of patterns of responses may be influenced by the number of moderators which is related, for example, to student enrollment. That is, units with many students and many lecturers may actually be easier to moderate due the available of potential co-assessors. All such identified case study limitations inform subsequent improvements to the RAMP process.

Preliminary feedback from unit coordinators, lecturers and university administration suggested strong support for mechanisms of continuous quality improvement in instructional processes, generally, and for the online RAMP process, specifically. Anecdotal comments revealed that unit coordinators appreciated automated approaches to record keeping due to ease of completion. Reportedly, unit coordinators shared their experiences with members of their instructional teams thereby rallying enhanced efforts to increase the validity of marks via assessment moderation practices. Lecturers mentioned that although increased expectations of assessment moderation were time consuming, celebration of their efforts also appeared to increase following implementation. University administrators eagerly embraced detailed information based on actual data which improved practice which was directly related to accreditation standards compliance. The very process of RAMP completion increased awareness of the importance and requirements of assessment moderation. In models of continuous quality improvement, record keeping highlights expectations of excellence which in turn leads to increased focus on achieving and surpassing minimum standards which in turn drives instructional excellence. The focus of RAMP was on ensuring the integrity of student marks and grades which is pivotal to maximizing student learning. University instructional staff responded positively to efforts to promote student learning and, in this regard comply with university policy. RAMP appeared to translate policy into visible and effective instructional practice.

References

- Adie, L. 2012. "Learning as Identity and Practice through Involvement in Online Moderation." *Educational Assessment, Evaluation and Accountability* 24: 43-56.
- Adie, L., Klenowski, V., and Wyatt-Smith, C. 2012. "Towards an Understanding of Teacher Judgment in the Context of Social Moderation." *Educational Review* 64 (2): 223-240.
- Adie, L., Lloyd, M., and Beutel, D. 2013. "Identifying Discourses of Moderation in Higher Education." *Assessment & Evaluation in Higher Education* 1-10.
- Annetts, S., Jones, U., and Deursen, R. V. 2013. "An Innovative Review of an Undergraduate Dissertation Double Marking Policy." *Innovations in Education and Teaching International* 50 (3): 308-317.
- Ashworth, M., Bloxham, S. and Pearce, L. 2010). Examining the Tension between Academic Standards and Inclusion for Disabled Students: The Impact on Marking of Individual Academics' Frameworks for Assessment. *Studies in Higher Education* 35 (2): 209-223.
- Bird, F. L. and Yucel, R. 2013. Improving Marking Reliability of Scientific Writing with the Developing Understanding of Assessment for Learning Programme. *Assessment & Evaluation in Higher Education* 38 (5): 536-553.
- Bloxham, S. 2009. "Marking and Moderation in the UK: Face Assumptions and Wasted Resources." *Assessment and Evaluation in Higher Education* 34 (2): 209-220.
- Bloxham, S. 2012. "You can see the Quality in Front of your Eyes: Grounding Academic Standards between Rationality and Interpretation." *Quality in Higher Education* 18 (2): 185-204.
- Bloxham, S., Boyd, P. & Orr, S. 2011. Mark my Words: The Role of Assessment Criteria in UK Higher Education Grading Practices. *Studies in Higher Education* 36 (6): 655-670.
- Buglear, J. 2011. "Grading and Academic Freedom: An English Academic's Angle on Hill's Contentious Triangle." *Quality in Higher Education* 17 (1): 101-104.
- Colton D. 2000. Quality Improvement in Health Care: Conceptual and Historical Foundations. *Evaluation and the Health Professions* 23:7-42.
- Dew, J. R. and McGowan Nearing, M. 2004. *Continuous Quality Improvement in Higher Education*. Westport, CT: Praeger.

- Gamble, I. 2011. *Continuous Improvement Models that Focus on Teaching and Learning*. Association of Independent Schools in South Australia. Accessed November 6, 2013. http://www.ais.sa.edu.au/__files/f/92190/AISSA%20-%20Continuous%20Improvement%20Models%20publication.pdf
- Goos, M. and Hughes, C. 2010. "An Investigation of the Confidence Levels of Course/Subject Coordinators in Undertaking Aspects of their Assessment Responsibilities." *Assessment & Evaluation in Higher Education* 35 (3): 315-324.
- Hartley, J. 2005. "Case Study Research." In *Essential Guide to Qualitative Methods in Organizational Research* edited by Catherine Cassell and Gillian Symore, 323-333 London: Sage
- Hunter, K. and Docherty, P. 2011. "Reducing Variation in the Assessment of Student Writing." *Assessment & Evaluation in Higher Education* 36 (1): 109-124.
- Hunter, S. B., Ober, A. J., Paddock, S. M., Hunt, P. E. and Levan, D. 2014. "Continuous Quality Improvement (CQI) in Addiction Treatment Settings: Design and Intervention Protocol of a Group Randomized Pilot Study." *Addiction Science & Clinical Practice* 9(4): 1-11.
- Johnson, S. 2013. On the Reliability of High-Stakes Teacher Assessment. *Research Papers in Education* 28 (1): 91-105.
- Klenowski, V. 2011. "Assessment for Learning in the Accountability Era: Queensland, Australia." *Studies in Educational Evaluation* 37 (1): 78-83.
- Klenowski, V. and Wyatt-Smith, C. M. 2010a. Standards, Teacher Judgment and Moderation in the Contexts of National Curriculum and Assessment Reform. *Assessment Matters* 2: 107-131.
- Klenowski, V. and Wyatt-Smith, C. M. 2010b. Standards-Driven Reform Years 1-10. *Australian Educational Researcher* 37 (2): 21-39.
- Kuzich, S., Groves, R., O'Hare, S. and Pelliccione, L. 2010. "Building Team Capacity: Sustaining Quality in Assessment and Moderation Practices in a Fully Online Unit." *ATN Assessment Conference 2010, University of Technology Sydney*, 1-10. Accessed November 6, 2013. <http://www.iml.uts.edu.au/pdfs/Kuzich.pdf>.
- Mahmud, S., Sanderson, G., Yeo, S., Briguglio, C., Wallace, M., Hukam-Singh, P. and Thuraingsam, T. 2010. "Moderation for Fair Assessment in Transactional Review." *Australian Learning & Teaching Council*, Accessed November 6, 2013. <http://altc-tne-moderation.wikispaces.com/file/view/ALTCLitReview6April2010.pdf>.

- Maxwell, G. S. 2010. "Moderation of Student Work by Teachers." *International Encyclopaedia of Education* 3: 457-463.
- Nuttall, M. 2007. "Moderation: Towards Greater Consistency in Marking." In *Developing Effective Assessment in Higher Education: A Practical Guide* edited by S. Bloxham and P. Boyd, 118-128. Maidenhead: Open University Press.
- Orr, S. 2010. "We kind of try to merge our own experience with the objectivity of the criteria: The role of Connoisseurship and Tacit Practice in Undergraduate Fine Art Assessment." *Art, Design & Communication in Higher Education* 9 (1): 5-19.
- Orr, S. and Bloxham, S. 2013. "Making Judgments about Students Making Work: Lecturers' Assessment Practices in Art and Design." *Arts and Humanities in Higher Education* 12 (2-3): 234-253.
- Qualtrics. 2014. *Online Survey Software*. Demonstration available at <http://qualtrics.com/>
- Rubenstein, L., Khodyakov, D., Hempel, S., Danz, M., Salem-Schatz, S., Foy, R., O'Neill, S., Dalal, S. and Shekelle, P. 2014. "How Can We Recognize Continuous Quality Improvement?" *International Journal of Quality Health Care* 26(1): 6-15
- Sadler, D. R. 2009. "Indeterminacy in the use of Preset Criteria for Assessment and Grading in Higher Education." *Assessment & Evaluation in Higher Education* 30 (2): 175-194.
- Sadler, D. R. 2013. "Assuring academic achievement standards: From moderation to calibration." *Assessment in Education: Principles, Policy & Practice*, 20(1): 5-19.
- Smith, C. 2012. "Why should we bother with Assessment Moderation?" *Nurse Education Today* 32 (6): 45-48.
- Singh, A. S. 2014. "Conducting Case Study Research in Non-Profit Organizations." *Qualitative Market Research: An International Journal* 17(1): 77-84.
- Thuraisingam, T., Singh, P. K., Sanderson, G., Mahmud, S., Yeo, S., Briguglio, C. and Wallace, M. 2010. "An Activity Theory Approach to Moderation for Fair Assessment in Transnational Education." Australian Learning & Teaching Council. Accessed November 6, 2013.
- http://espace.library.curtin.edu.au/webclient/StreamGate?folder_id=0&dvs=1378900662714~583&usePid1=true&usePid2=true.

- Van der Schaaf, M., Baartmanb, L. and Prins, F. 2012. "Exploring the Role of Assessment Criteria during Teachers' Collaborative Judgment Processes of Students' Portfolios." *Assessment & Evaluation in Higher Education* 37 (7): 847–860.
- Wallace, M., Mahmud, S., Sanderson, G., Yeo, S., Briguglio, C., and Hukam-Singh, P. 2010. "Moderation of Assessment in Transnational Higher Education." In *Collaborative Approaches to Moderation of Assessment in Transnational Education* edited by Gavin Sanderson, 1-11. Melbourne: RMIT University.
- Wyatt-Smith, C., Klenowski, V. and Gunn, S. 2010. "The Centrality of Teachers' Judgment Practice in Assessment: A Study of Standards in Moderation." *Assessment in Education: Principles, Policy & Practice* 17 (1): 59-75.

Figure 1

Record of Assessment Moderation Practice (RAMP) Interface: Assessment Description

ASSESSMENT 1 - PERCENTAGE AND TYPE

Percentage (weighting)
(No % symbol, insert numbers only)

Select the assessment type from the following:

- Test, quiz, exam
- Presentation
- Written submission

Figure 2

Record of Assessment Moderation Practice (RAMP) Interface: Textbox Explanations

Assessment 1 - Moderation

Has this assessment changed since it was last taught to prevent plagiarism?

No

Yes

How has this assessment changed since it was last taught?

Is this assessment moderated?

Yes

No

Why isn't this assessment moderated?

Figure 3

Record of Assessment Moderation Practice (RAMP) Interface: Response-Options

Select all statements that apply to this assessment:

- Moderators discuss the information provided to students that describes the assessment
- Moderators discuss the marking rubric with each other
- Moderators review and discuss samples of marked assessments with completed rubrics (e.g. fail, credit, distinction, high distinction from previous study periods)
- During the marking process, moderators provide opportunities for co-assessors to discuss and validate their evaluation of student assessments (e.g. marked assessments in a file exchange program)
- Moderators co-assess all failed assessments
- Moderators refer all assessments that cannot be moderated successfully within the unit to the Quality of Assessment Panel
- Moderators adhere to a single release of student feedback under the direction of the Unit Coordinator

Select all statements that apply to this assessment:

- This assessment is aligned with the unit learning outcomes and syllabus
- This is a new assessment
- This assessment is offered in more than one location
- This is a centrally scheduled examination
- In the case of repeat assignments/examinations, this assessment task distinguishes itself from those used in previous semesters
- A pre-marking meeting clarifies marking criteria and feedback practices
- The requirements of the assessment task and marking and feedback practices are explained to students
- Students provide feedback to peers prior to final submission for summative marking
- Marking checks are undertaken to confirm marking integrity

Select all statements that apply to this assessment:

- Marking guides/rubrics are used to minimise variation across the marking team and to assist individual markers with consistency when marking multiple assessments
- Rotational marking is used where the assessment is divided into sections and each section is marked by the same marker
- In the case of a large number of assessments marked by the same person, a sample of papers marked early in the process are reviewed by the marker for consistency
- Double marking is undertaken, where a random selection of assessments are independently assessed and by two (or more) markers and results compared

Figure 4

Record of Assessment Moderation Practice (RAMP) Data Output: Type of Assessment

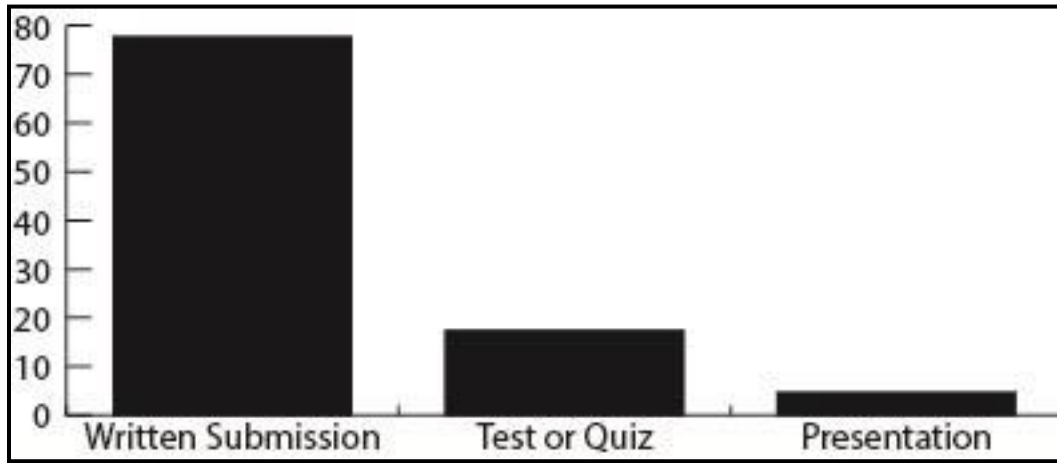


Figure 5

Percentage of Unit Coordinators Reporting Specific Assessment Moderation and Integrity Strategies

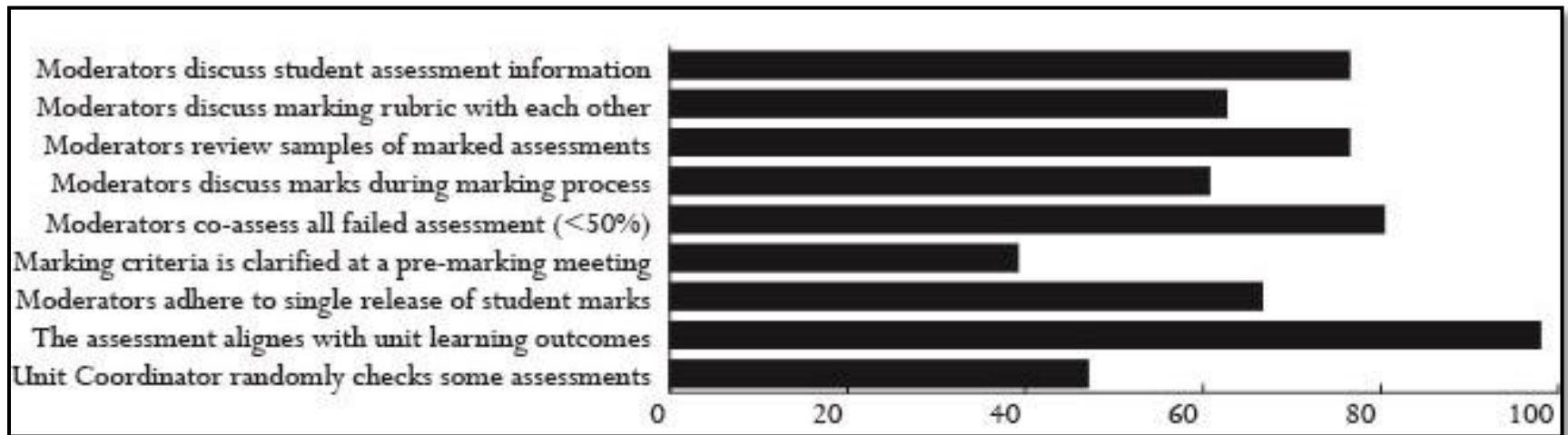


Figure 6

Percentage of Units Reporting Number of Moderation and Integrity Strategies for First Written Assessment

