Good Practice in Assessment Case Study – School of Public Health: Assuring Academic Achievement Standards and investigating grade inflation
Royce Sadler¹, Lynda Davies¹, Roger Moni¹, Don Stewart², Claire Brown².

¹ The Griffith Institute for Higher Education
² School of Public Health

This case study demonstrates how standards of academic achievement can be assured at a whole-of-school level, providing confidence to both staff and students that the grades issued represent the level of student achievement as evidenced in the quality of the work they have produced.

Context
The question of how to assure academic achievement standards is being asked across the higher education sector globally. The OECD launched an international project AHELO to investigate and report on ways in which academic standards manifest themselves in different degrees from different countries and to find a way to compare them. In Australia, there a number of approaches under consideration. The new Tertiary Education Quality Standards Agency (TEQSA) will almost certainly look at how universities maintain and monitor standards; Australian Universities Quality Agency (AUQA) already keeps a close watch on university processes. The Australian Learning and Teaching Council (ALTC) is in the process of replicating in Australia the UK’s Quality Assurance Agency (QAA)’s decade-long quest to establish sets of explicitly articulated discipline standards, and this project is now well underway. Professional and industry accrediting bodies require particular elements to be taught in particular ways within many degrees as part of their drive to ensure their expectations and needs are met. University groups such as the Australian Technology Network (ATN) are developing frameworks against which their members will benchmark themselves.

In this context, and partly in response to AUQA, Griffith launched an assessment reform programme at the heart of which lie the concepts of assuring academic standards and consensus moderation. The School of Public Health decided to review its assessment practices and explore ways in which it could more consistently and confidently assure its academic achievement standards. It consequently commissioned the Griffith Institute for Higher Education (GIHE) to assist in, and facilitate, that investigation.

A way forward
In April 2010 the project team designed a proposal in conjunction with the Head of School and presented it at a whole-of-school workshop for consideration. It was agreed to explore an approach to assuring academic achievement standards at a whole-of-school level and investigate whether the School was experiencing grade inflation by looking at the possible contributors within the course assessment plans and methods. The proposal included three components:
(a) build a shared understanding of the “Five Core Concepts of Assessment” through whole-of-school workshops led by Royce Sadler;
(b) participate in masterclass colloquia at which school members would work in small groups to read and discuss two articles “Grade Integrity” and “Fidelity” by Royce Sadler; and
(c) engage in panel activities focusing on three critical elements of assessment quality: academic achievement; assessment task quality; and consensus moderation.

Diagram 1. Phases of the project

Throughout 2010 the School engaged in all three elements – the substance of which is described below.

About the School workshops
With Head of School sponsorship, the project team was invited to work with staff at whole-of-school meetings and the school’s learning and teaching forum at key stages of the project’s cycle. Particular attention was paid to building a shared understanding of core concepts of assessment. This phase was led by Royce Sadler based on his work over the last three decades. These concepts of assessment formed the framework upon
which the different activities of the project were based. A brief summary of those concepts is provided below, however, please see the references at the end of the case study for details of the papers in which they are explained in full.

**Academic Achievement**
Summative assessment must only give credit for true achievements, not non-achievements. A ‘pure’ object (student work) is one for which the quality is determined by analysis followed by a decision or judgement and coded as part of the marking process. That code is in turn converted into a grade. The trueness (of the achievement) is termed ‘fidelity’ and is a precondition for maintaining grade integrity. There are various contaminating non-achievement components of a grade that are commonly used to leverage student behaviour such as marks for attendance or effort. Other contaminants include student works that have not addressed the task sufficiently to let the marker know what the quality of the work was like. When asking for a “pineapple” teachers are sometimes given a “boot”. That means there is no way of judging whether the “boot” was a good, bad or middling “pineapple”. The student response needs to address the task specifications well enough that it can be judged as a “pineapple”. For fidelity in grading to be maintained, contaminants such as marks for attendance and submission of “boots” cannot be considered. Other ways have to be sought to shape student behaviour without adversely affecting the integrity of the grade. (See Panel 1 for activity based on this concept.)

**Commensurability**
Commensurability is the correspondence between the mark or grade awarded and the level of achievement. Quality can be thought of as existing on a continuum with extremely poor at the low end and excellent at the other. The grade must correspond to quality (wherever it sits on that continuum) and the grading scale used should reflect the full range of possibilities. Of special importance are the thresholds for PASS and HIGH DISTINCTION. (This concept influences the judgements and discussions in Panel 3.)

**Comparability**
Things which are unlike may nevertheless be comparable; this has two aspects. The first meaning of comparable could be written (and pronounced) as ‘compare-able’ meaning able to be compared, but always with respect to one or more criteria. The second meaning may be written (and pronounced) as ‘com-prable’, meaning of similar worth, merit or value. Things cannot be com-prable unless they are first of all compare-able. Comparability of grading standards across courses is the key issue that Assessment Boards always try to grapple with. Their main method to date has been to arrange for adjustments to grade distributions, to ‘make them com-prable’. The current assessment reform involves working to solve the comparability problem (a) substantively, and (b) at its source, where grading decisions are made. This is quite a different exercise from re-adjusting grades, and requires access to student works from at least two courses, and direct judgments about their comparability.
Standards-referencing
Standards-referencing is the **grading principle** which grades student works according to established standards. To the best of our ability, we should mark or grade student work strictly according to its quality – in absolute rather than relative terms. Ideally, the ‘standards’ should be fixed, and remain that way year after year, regardless of cohort. This is a principle Griffith University has espoused since it was founded in the early 1970s. (This concept influences the judgements and discussions in Panel 3.)

Consensus Moderation
Consensus moderation is the main tool for attaining grade integrity in practice. It is already widely used when large numbers of responses to a single assessment task are evaluated by a team of markers. Team members each mark a sample of responses; they then confer, debate and agree on the standards for that event; then go ahead with the rest of the marking. This is simply a particular form of peer review, which is the method endorsed for evaluating research grant proposals and academic manuscripts for publication. When appropriately scaled up, consensus moderation can address all of the issues required for grade integrity. Applied to academic achievement standards, it involves clusters of lecturers in ‘cognate’ courses setting comparable standards for use in those courses. The ownership of the process and the results remains with the academics who design and teach courses.

About the masterclass colloquia
Academic staff members worked in small groups to read and discuss the key concepts put forward in two of Royce Sadler’s papers. A short abstract from each of the papers is provided below, however, please see the references at the end of the case study for access to the papers in full. Prompt questions were:

- *How did you react to the articles?*
- *What parts don’t you understand or have problems with?*
- *Are there ideas that don’t seem to connect/make sense?*

“Grade Integrity”: “In this article, grade integrity is defined as to the extent to which each grade awarded (either at the conclusion of a course or module of study or for an extended response to an assessment task), is strictly commensurate with the quality, breadth and depth of a students’ performance. The three basic requirements for this aspiration to be realised are, in order: assessment evidence of a logically legitimate type; evidence of sufficient scope and soundness to allow for a strong inference to be drawn; and a grading principle that is theoretically appropriate for coding the level of a student’s performance. When further developed, the general approach outlined could produce positive side benefits, including ways of dealing with grade inflation.” (Sadler, 2009.)
“Fidelity”: “If a grade is to be trusted as an authentic representation of a student’s level of academic achievement, one of the requirements is that all the elements that contribute to that grade must qualify as achievement, and not be something else. The implications of taking this proposition literally turn out to be far reaching. Many elements that are technically non-achievements are routinely incorporated into grades and thereby act as contaminants. A variety of credits and penalties are often included with the intention of helping shape student behaviours or improve their learning. Reversing the situation has ramifications not only for assessment and grading practices but also for the ways in which curriculum and teaching are conceptualised, designed and engaged in.” (Sadler, 2010)

Between 5 July and 4 August 2010, eleven staff members in their groups worked with the author, talking through various issues, challenging and resolving ideas as they apply to the context in which PBH works. After the sessions, fifteen main comments from participants were extracted from notes taken during masterclass colloquia. These comments were then matched to one or more of the five core concepts of assessment to identify initial, salient ideas around assessment in PBH (Table 1). Unsurprisingly, many comments could be matched to two concepts. However, most responses reflected the importance of academic achievement.

Table 1. Matching responses of masterclass participants to core concepts of assessment in PBH

<table>
<thead>
<tr>
<th>Core concept</th>
<th>% of comments</th>
<th>Illustrative examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic Achievement</td>
<td>45</td>
<td>Some first-year students think they already know what ‘quality’ of achievement means (some think attendance should contribute to grades), so these students need special assistance to learn what we value.</td>
</tr>
<tr>
<td>Commensurability</td>
<td>5</td>
<td>Grades are inflated for some courses and I’m revising assessment for these courses.</td>
</tr>
<tr>
<td>Comparability</td>
<td>10</td>
<td>‘The bar is all over the place’, regarding the large differences across both programs and universities – standards are all over the place...but how do academics (with say 2 versus 20 years of experience) know where to set the bar?</td>
</tr>
<tr>
<td>Standards-referencing</td>
<td>25</td>
<td>We should not be judging first-year students on their previous (academic) performances although different universities get different students (in terms of academic achievements).</td>
</tr>
<tr>
<td>Consensus Moderation</td>
<td>15</td>
<td>The paper makes sense theoretically but I want to look at applications of ‘what we mean by achievement’ across the whole school, not just as an individual academic.</td>
</tr>
</tbody>
</table>
About the panel activities

Three panel activities were designed to give staff an opportunity to consider the different aspects of assuring academic standards as part of the day-to-day assessment enterprise and what happens when (or if) those aspects are not implemented at a whole-of-school level. School staff selected which activity they wanted to undertake. Two definitions pertinent to Panel activities 1 and 2 were provided as part of the resources.

Learning:
“For my money, learners can be said to have learned something when three conditions are satisfied. They must be able to do, on demand, something they could not do before. They have to be able to do it independently of particular others, those others being primarily the teacher and members of a learning group (if any). And they must be able to do it well. Assessment of learning should be directed towards gathering evidence for drawing inferences about capability under these conditions, not the scaffolded conditions” (Sadler “Perils”, 2007).

Academic Achievement:
“...is taken as the attainment of an identifiable level of knowledge or skill [proficiency, capability, competence or performance] as determined through evaluating performances on assessment tasks, or through observation of relevant behaviours in specified settings” (Sadler, “Fidelity”, 2009, 4).

Examples of common non-achievement inclusions for a grade (often used as leverage mechanisms to shape student behaviours) are attendance, participation, or effort (Sadler “Grade Integrity”, 2009, 809-21). These short definitions acted as a scaffold to the discussions, reminding members of earlier conceptual work they had participated in during the workshops and masterclasses on what learning and academic achievement are and are not.

Panel 1: What is Academic Achievement?

Data reviewed by members were de-identified assessment plans and learning outcomes from a cognate set of existing School courses with enrolments over 10. Members were asked to consider the question “How is academic achievement demonstrated in PBH courses?”

From each assessment plan, members identified which tasks assessed academic achievement (marking them as ‘AA”) and which tasks assessed non-academic achievement (marked as “NA”). On a worksheet provided, members wrote the reasons for their decisions, then shared and talked about them, justifying their judgements. Members reached a consensus across the group, recording any changes to their initial decisions on the worksheet.
The next set of questions guiding discussion was based around managing any consequences of reaching a group understanding of how academic achievement was tested in the selected assessment plans and whether there needs to be changes at a whole-of-school level.

Members were asked to reconsider the assessment plans and those tasks identified as non-academic achievement and discuss them in relation to the following prompts:

a) What functions are the non-achievement tasks serving?

b) Can we “comb them out”?

c) What would we replace the non-achievement tasks with if the function needs to remain?

d) How do we manage student reactions?

Panel 2: Assessment Task Quality

To provide authentic primary data from which to work, examples of assessment tasks were selected, de-identified and copied from a cognate set of existing School courses (across various year-levels) with enrolments over 10. Members of the panel were asked to peruse the assessment task specifications provided and the learning outcomes published in the course profile. They were asked to consider the following question “Are the tasks of high quality?” and judge the quality of each assessment task specification on a high/medium/low scale in relation to the following characteristics:

a) To what extent does it require students to demonstrate achievement with a particular focus on higher-order thinking?

b) Would student responses produce evidence of the right types of learning (or academic achievement) in PBH?

c) Can the task specifications be taken literally?

Members recorded their judgements on a sheet provided for the activity, then talked about and shared those judgements in order to reach group consensus by calibrating their judgements with their colleagues’. If any member decided to change their initial judgement, they recorded that change accordingly. Finally, members were asked to note “what steps (if any) could the convenors of these courses take to improve these tasks?”

Panel 3: Consensus Moderation

First, members of the panel were asked to put aside concerns of attrition, grading proportions and read a common set of de-identified student scripts (obtained with student permission). Second, members followed the Praise, Question, Polish protocol for judging work. They wrote comments on the scripts in response to the following prompts:

(a) “What are the strengths of this piece of student work? (Praise – broad aspects, make constructive comments)”;
(b) “What are its weaknesses? (Question – write questions to students, asking them to clarify unclear aspects)”; and

(c) “What would you recommend students could do to improve their work? (Polish)”. Third, members marked the student scripts independently using the High Distinction (D), Distinction (D), Credit (C), Pass (P), or Fail (F) coding. An important aspect of this process was for them to record justification of their marks/comments to assist in the next stage of the activity. The fourth stage required members to talk and share their judgements and marks. The purpose of this discussion was to reach a shared understanding of each others’ expectations of quality and to articulate what they looked for in an HD, D, C, P or F piece of work in relation to the specific scripts under consideration. Finally, group consensus was achieved as they calibrated their judgements and agreed on the mark they would “award” the pieces of work. If in this process of calibrating judgements, any change to their original decision occurred, they recorded such change and the reasons for it. (See Appendix B for summary of outcomes of panel activities)

Demonstration of Benefits

The project was designed to be a ‘whole-School’ activity with buy-in from all levels of academics across all programs. Assessment was re-framed as not simply an individual identification of what was perceived to be appropriate learning on the part of the student and an assessment of the extent to which the student had achieved mastery of that learning, but that there is also a School/Program context of learning. We needed to have some coherence and continuity above and beyond the individual academic/student relationship. This Project facilitated a re-focusing at this higher level.

In addition, the Project provided a real-world response to increasing pressure to improve student evaluations of learning and teaching. With a University-wide formula that emphasizes student assessment scores of courses and teaching as a critical component for decisions about academic resources, the School needed to do more to address areas of weakness. This Project focused on learning and teaching and provided an academically legitimate and logical opportunity to improve assessment of student learning.

The process as outlined above has led many of the School academic staff to reflect on the reasons for their teaching, what outcomes can be expected by students, and the ways in which achievement of learning outcomes can be assessed. Starting from a position of dealing with apparent grade creep, the School is now grappling with the broader questions involving the purposes and grading of assessment items.

Principles of Good Practice illustrated in this project

As part of the national ALTC Teaching Quality Indicators Project (TQI), Griffith developed a set of quality indicators for assessment. The Statements and Quality Indicators of Good Practice in Assessment have been disseminated nationally and used by universities across the sector. The indicators have informed the review of Griffith’s assessment policy and underpin the work on the Promoting Good Practices in Assessment Project.
While the initiative in PBH is clearly related to the international and national exploration of improving ways to assure academic achievement standards, it can also be situated within the context of the TQI project.

It is useful, therefore, to reproduce the Principle upon which the *Statements and Quality Indicators* were constructed, and to identify the Statements and Quality Indicators relevant to this case study.

**Principle**

Griffith University’s Teaching Quality Indicators Project has been guided by research into the theoretical literature on good principles and practices of assessment in higher education; assessment practice at Griffith University; and assessment policies used across Australia and overseas.

This work has shown that assessment inevitably shapes how students approach learning, including what they focus on and how they go about learning it, and is used for a variety of purposes. Necessarily, assessment underpins the core values and principles of the University’s learning and teaching strategic plans and a clear enunciation of what drives assessment at the University is important for students, staff, and the broader community.

It is accepted, therefore, that the primary purpose of assessment is to:

- promote student learning; and
- provide information upon which judgements are made about students’ work and the standards their performances exhibit.

It is clear that the Head of School, PBH believed that the existing assessment methods and practices in relation to academic achievement standards needed investigation and as such he began a process of review that looked at whole-of-school understandings of grade integrity and whether contributors to grade inflation could be found in current practices.

Such an approach is reflected in the following Statement of Good Practice and shows the link between the stages of review and the subsequent improvement sought through the professional development activities in the workshops and masterclasses.

**Statement of Good Practice #4: Assessment policies and practices are planned, implemented, reviewed and improved**

**This occurs when:**

4.1. assessment practices are given consideration in cyclical reviews of teaching, Programs, Courses and academic units;

4.2. staff use feedback from peers and students to improve subsequent assessment practices;
4.3. professional development opportunities are provided to staff to assist them in improving assessment practices;

4.4. staff participate in professional development activities relevant to assessment (e.g. workshops, conferences, relevant literature etc.,)

The design of the Panel activities tackled other aspects of assessment practice.

Panel 1: “What is academic achievement?” interrogated the issues around what academic achievement looks like in PBH programs and courses, and what are the academic achievement standards expected of PBH students.

Panel 2: “Assessment Task Quality” reviewed a selected set of assessment task specifications and course learning outcomes to determine whether they: adequately require PBH students to focus on higher-order thinking; produce evidence of the appropriate types of learning and levels of academic achievement; can be taken literally.

Panel 3: “Consensus Moderation” members had to judge a set of student scripts and independently determine what grade they would have awarded to the works. During discussions about their judgements, members reached consensus on the quality of the academic achievements demonstrated in the works, developing a shared understanding of the expectations each held in relation to quality. They consequently calibrated their judgements for future marking tasks.

The activities of all three panels are reflected in Statements #1 and #3 respectively.

**Statement of Good Practice #1: Assessment tasks are designed to advance student learning**

**This occurs when:**

1.1. there is consistency between what is assessed, course aims and objectives, graduate attributes, and the way things are taught;

1.3. tasks test appropriately the increasing complexity of intellectual activity, and require students to demonstrate their growth in understanding and development of skills;

**Statement of Good Practice #3: Assessment practices are fair, produce marks and grades that are reliable and valid, and certify students’ achievements.**

**This occurs when:**

3.1. the mark allocated for an assessment task reflects the standard of the students’ work regardless of how other students perform;

3.2. mark allocation, and the system of marking, properly represents the overall judgement of the standard achieved by the students’ performance;
3.3. benchmarking occurs at School, Faculty or Group level to calibrate academics’ understanding of the academic achievement standards used when judging the quality of students’ work; and

3.4. moderation of assessment tasks occurs between members of teaching teams or relevant School colleagues.

The project’s focus on assuring academic achievement standards within PBH has built staff capacity and a shared understanding of the issues involved. Consequently it has provided a foundation for future work that will strengthen the School’s approach to assessment reform.

Future directions
The work conducted throughout 2010 piloted various activities that together explored an approach to achieving consensus moderation. The School as a collective will be meeting in November to work through the outcomes of those activities and determine what will be adopted for wider implementation in 2011 and beyond. Accordingly it is anticipated that an Action Plan will be drawn up to assist the School in its continued work in the area of assessment review and consensus moderation. Each of the Panels in 2010 recommended some possible actions which will be considered at the November workshop. They are summarised below.

Panel 1: (a) Clarify across PBH how to assess and manage group work. (b) Establish shared understanding of what quality of written work is not acceptable and communicate this to students.

Panel 2: (a) Shared repository of assessment tasks across PBH would assist other experienced staff to understand standards expected. (b) 1-2 times per year share innovations in assessment across PBH. (c) Could have brief document (‘kit’) summarising PBH assessment rules to guide staff (and students?).

Panel 3: (a) Run this exercise across whole School. Maybe all year-level teachers work together in first instance judging their relevant year-level work, and then run another with members who teach in any combination of first-year to postgraduate so that they get a sense of quality of work from students as they progress through the program. This horizontal and vertical arrangement would give the whole School a clear picture of what quality looks like, is expected and needed in progressive years.

Acknowledgements
The project team (Royce Sadler, Roger Moni, Lynda Davies, GIHE; and Don Stewart and Claire Brown, PBH) wishes to thank Jo-Anne Clark (SRA) and Duncan Nulty (GPA Project Leader) for their assistance. The engagement of the School of Public Health (PBH) staff has been integral to the success of this project and without their commitment it would not have been possible. In alphabetical order they are: Helen Cooper, Ben Desbrow, Ian Edwards, Debby Findlay, Neil Harris, Chris Irwin, Vanessa
Lee, Kathleen Lilley, Michael Leveritt, Zoe Murray, Michelle Palmer, Allen Ross, Fiona Rowe, Bernadette Sebar, Shawn Somerset, Mark Storrs, and Jing Sun. In particular, the project team acknowledges the leadership shown by the members of the Panel activities.

References


### Appendix A

<table>
<thead>
<tr>
<th>Phase</th>
<th>Timeline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial scoping discussions and opening workshop</td>
<td>Aug 2009</td>
</tr>
<tr>
<td>Auditing course outlines; compiling data on assessment task types and weightings for all PBH undergraduate courses with enrolments over 10</td>
<td>Nov 2009 – Dec 2009</td>
</tr>
<tr>
<td>PBH school workshops (6)</td>
<td>Feb – Sept 2010</td>
</tr>
<tr>
<td>Masterclasses with PBH academic staff (4)</td>
<td>Aug 2010</td>
</tr>
<tr>
<td>Pilot improvement strategy: panel activities (3)</td>
<td>Sept - Oct 2010</td>
</tr>
<tr>
<td>“Bringing it all together” and develop action plan for implementation school workshops (2)</td>
<td>Oct – Nov 2010</td>
</tr>
<tr>
<td>Implement: Assuring Academic Achievement Standards</td>
<td>2011</td>
</tr>
</tbody>
</table>
Appendix B

Bringing it all together
School of Public Health and Assuring Academic Achievement Standards
(Investigating Grade Inflation)

School Workshops – six conducted over period September 2009 - September 2010

Five Core Concepts of Assessment

(1) **Academic achievement**: ‘Fidelity’ as a precondition for
(2) **Commensurability**: Correspondence between mark/grade awarded and achievement
(3) **Comparability**: Things unlike can be comparable
(4) **Grading principle**: Grading according to established standards
(5) **Consensus moderation**: The main tool for attaining grade integrity

Masterclasses – conducted over period August 2010

**Learning:**

“For my money, learners can be said to have learned something when three conditions are satisfied. They must be able to do, on demand, something they could not do before. They have to be able to do it independently of particular others, those others being primarily the teacher and members of a learning group (if any). And they must be able to do it well. Assessment of learning should be directed towards gathering evidence for drawing inferences about capability under these conditions, not the scaffolded conditions.”

(Sadler “Perils”, 2007)

**Academic Achievement:**

“...is taken as the attainment of an identifiable level of knowledge or skill [proficiency, capability, competence or performance] as determined through evaluating performances on assessment tasks, or through observation of relevant behaviours in specified settings.”

(Sadler, “Fidelity”, 2009, 4)

*Common non-achievement inclusions (often used as leverage mechanisms to shape student behaviours) e.g., attendance, participation, effort.*

(Sadler “Grade Integrity”, 2009, 809-21)

Panel Activities – 3 Panels conducted over period September – October 2010

Panel 1: What is academic Achievement?
Panel 2: Assessment Task Quality
Panel 3: Consensus Moderation
Panel 1: What is Academic Achievement? (28 September 2010)

Summary outcome

Panellists: Claire Brown, Michelle Palmer, Fiona Rowe, Bernadette Sebar, Mark Storrs

<table>
<thead>
<tr>
<th>Activity</th>
<th>Judgement</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Academic Achievement (AA)</td>
<td>Non-academic Achievement (NA)</td>
</tr>
<tr>
<td>Example 1.1</td>
<td>Example 1.1</td>
<td></td>
</tr>
<tr>
<td>Task 1</td>
<td>5 panellists agree NA</td>
<td></td>
</tr>
<tr>
<td>Task 2</td>
<td>5 panellists agree AA</td>
<td></td>
</tr>
<tr>
<td>Task 3</td>
<td>3 panellists judge AA</td>
<td>Task 3: Either change task to reflect AA for individual students, and/or change its weighting</td>
</tr>
<tr>
<td></td>
<td>2 panellists vary NA</td>
<td></td>
</tr>
<tr>
<td>Task 4</td>
<td>5 panellists agree AA</td>
<td></td>
</tr>
<tr>
<td>Example 1.2</td>
<td>Example 1.2</td>
<td></td>
</tr>
<tr>
<td>Task 1</td>
<td>5 panellists agree NA</td>
<td></td>
</tr>
<tr>
<td>Task 2</td>
<td>5 panellists agree AA</td>
<td></td>
</tr>
<tr>
<td>Task 3</td>
<td>3 panellists saw mixture of both AA &amp; NA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 panellist judged NA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 panellist AA</td>
<td></td>
</tr>
</tbody>
</table>

Major Recommendations: (a) Clarify across PBH how to assess and manage group work. This process will combine both conceptual understanding of purpose of group work and managing the consequences of decisions. Examples cited by panellists: some don’t assess group work; some assess individual contributions to group work; some use non-graded pass or satisfactory/unsatisfactory, some give whole-of-group mark. Method will depend on course content and type. Tied up in this is the question of how to assess generic skills if those skills are not part of core content. (b) Establish shared understanding of what quality of written work is not acceptable and communicate this to students.
**Panel 2: Assessment Task Quality** (12 October 2010)

**Summary outcome**

*Panellists: Helen Cooper, Zoe Murray*

<table>
<thead>
<tr>
<th>Activity</th>
<th>Judgement High, Medium, Low</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment Task 2.1</td>
<td>2 panellists agree M, though shifted from “H” and “M-H”</td>
<td>Might - provide students with larger data set; require tables or graphs that reflect work needs; 40% for language is high (across two criteria)</td>
</tr>
<tr>
<td>Assessment Task 2.2</td>
<td>2 panellists agree “L-M”</td>
<td>High-order thinking not required; 11 learning outcomes very specific; link between the biology and public health not clear; questions required recall</td>
</tr>
<tr>
<td>Assessment Task 2.3</td>
<td>2 panellists agree “L-M”, though only after negotiation</td>
<td>but one tended to “low-medium”, the other “medium-high”; the need for high-order thinking could have been more explicitly specified; criteria re-named by panellists; wording of the question was a little ambiguous; could test basic knowledge with on-line quizzes</td>
</tr>
</tbody>
</table>

**Findings:** Panellists were able to reach consensus around judgements of three assessment tasks. Task modifications were suggested e.g. clarifying the wording of tasks, and making them more challenging by stronger applications to public health.

**Major Recommendations:**
(a) Shared repository of assessment tasks across PBH would assist other experienced staff to understand standards expected. (b) 1-2 times per year share innovations in assessment across PBH. (c) Could have brief document (‘kit’) summarising PBH assessment rules to guide new staff.
**Panel 3: Consensus Moderation** (23 September 2010)

**Summary outcome**

*Panellists: Vanessa Lee, Allen Ross, Jing Sun*

<table>
<thead>
<tr>
<th>Activity</th>
<th>Judgement</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fail, Pass, Credit, Distinction, High Distinction</td>
<td></td>
</tr>
</tbody>
</table>
| Student Script A | 2 panellists gave Credit  
1 panellist gave Pass            | Reached consensus all panellists gave Credit                              |
| Student Script B | 1 panellist gave Distinction  
1 panellist gave Pass  
1 panellist gave Credit | Reached consensus all panellists gave Credit                              |
| Student Script C | 2 panellists gave Pass  
1 panellist gave Fail           | Reached consensus with two panellists agreeing to Pass (but with learning assistance to student and conditions for their future work). One panellist maintained judgement that it was not good enough to pass. |

**Findings:** Going beyond differences in expertise of panellists and acknowledgement of differences in how student learning is supported, staff hold different opinions of academic achievement standards in this form of academic writing (1000 word essay). Despite the differences there are common aspects too, and the judgements can be communicated and explored with shared understanding developed so that consensus moderation results in calibration of academics’ judgement over time.

**Major Recommendation:** (a) Run this exercise across whole School. Maybe all year-level teachers work together in first instance judging their relevant year-level work, and then run another with members who teach in any combination of first-year to postgraduate so that they get a sense of quality of work from students as they progress through the program. This horizontal and vertical arrangement would give the whole School a clear picture of what quality looks like, is expected and needed in progressive years.