

Moderation without Tears

Tony Cleford, Director of Design Practice

School of Architecture & Construction, University of Greenwich

Amanda Hopkins, Joint First Year Coordinator, BA Arts & Environment

Central Saint Martins College of Arts & Design

Abstract

Moderation is essential in academic assessment, but it is more difficult - and more necessary - in the assessment of work involving creativity and design.

In many schools a wide range of staff tutor the same project, bringing their different perspectives and values. They are often part-timers or visiting tutors, who may not know of earlier school-wide debates about the curriculum or project content. They may not share values held in common by full-time members of staff.

Yet these diverging perspectives - and the range of wildly different work they help produce - must be reconciled at assessment. It is those full time staff who often have the job of exploring, explaining, reconciling and resolving those differences. This case study provides an example of a mechanism for making the process of moderation easier.

Keywords: Assessment, Moderation, Design Teaching, Crit

Introduction

Moderation is essential in academic assessment, but it is more difficult - and more necessary - in the assessment of work involving creativity and design.

In many schools a wide range of staff tutor the same project, bringing their different perspectives and values. They are often part-timers or visiting tutors, who may not know of earlier school-wide debates about the curriculum or project content. They may not share values held in common by full-time members of staff.

Yet these diverging perspectives - and the range of wildly different work they help produce - must be reconciled at assessment. It is those full time staff who often have the job of exploring, explaining, reconciling and resolving those differences. This study's suggestions can make assessment and moderation easier.

The Problem

Typical assessment events

In many design-based disciplines, student work is assessed by crit (also called reviews, juries or pin-ups) and by the review of a student's oeuvre over a term, semester or year - the portfolio review. Often the student cohort is divided into tutorial groups, units or ateliers. These may be run by an individual, or sometimes a pair of tutors.

In the following, we give an example, describing a typical crit and portfolio review. Do any aspects seem familiar?

A typical crit?

Crits commonly involve tutors marking work by their own students and by students of other tutors. This is a useful piece of initial moderation in its own right: it brings a different perspective to bear.

However, if there is no mutual understanding of that perspective then there will be no agreement on why this work is better or worse. Without such a consensus, staff and students are likely to become frustrated and dissatisfied and moderation will be time consuming and inefficient.

The time issue itself is often contentious, causing confusion and delay. Sometimes tutors fail to agree on how long should be spent on each student's work. Some tutors argue that more interesting work should have longer spent on it. At assessment events, equal time slots are most appropriate, but the important thing is that they should be agreed by all the tutors before the day starts.

Having agreed a timetable, tutors may not stick to it, or the time scale may not be conveyed to the presenting students. As a result, as the day progresses, timing becomes undisciplined and students wait patiently as crits fall further behind schedule. The tutors' last review of the

day scheduled for 5pm is pushed back to 6, then to 7pm. Finally it is cancelled altogether as some tutors leave while others haven't even finished.

The academic in charge of collecting and collating marks from these tutors has little choice other than ask tutors to leave their marks in a pigeon hole before they turn the lights out.

The moderation meeting has to be reconvened the following week when everyone is in again, delaying marking and student feedback, or held sooner with fewer attendees – which brings inevitable complaints from the absentees about being kept out of the process.

A typical portfolio review?

If tutors are put in smaller groups, each reviewing some portfolios, it is understandable that everyone tends to want to see the best portfolios and also work in the 'twilight zone' between high fail and low pass. Often, especially where degree classifications hang on the results, all-group discussions extend to portfolios on the borderlines between classifications.

In our experience this often leads to discussions such as Tutor A arguing strongly for the reclassification of one of her students' portfolios. She swings Tutor B behind her, and then tutor C agrees, creating a critical mass of opinion. Once all staff are convinced, or have at least agreed to accept the changes, tutor B comments that because of the reassessment he thinks two of his students are in the same category and should be re-examined. Tutor C says if that is the case, he has another three students who now look miss-marked. The pile of portfolios in the "Done" pile starts to shrink as more and more earlier decisions become unfixed. It can be a long process.

And at the end of it, the chances are there will still be tutors who are dissatisfied with both the moderating process and the results it produced - and dissatisfied tutors often lead to dissatisfied students.

Examples

This example is derived from a school of architecture operating the unit, or atelier with two tutors per atelier. However the methods could be applied to other teaching systems and environments with minor modifications.

Assume as an example, that there are six ateliers (A-F), each staffed by two tutors:

Atelier A: tutors Anna and Agnetha

Atelier B: tutors Bjorn and Bennie

Atelier C: tutors Clara and Chrissie

Atelier D: tutors Dave and Duke

Atelier E: tutors Ed and Edie

Atelier F: tutors Fran and Francis

For the purposes of this example, assume each atelier has 20 students: we will call them A1-A20 in atelier A; B1-B20 in Atelier B and so on.

A 'How not to' example

At the crit, paired juries are set up. Anna and Clara spend the day each reviewing half their students together. They see students A1-A10 and C1-C10. At the same time, Agnetha and Chrissie review the other half: A11-A20 and C11-C20.

At the same time, Bjorn and Duke review half their students together (B1-B10, D1-D10) while Bennie and Dave review the other halves of their ateliers (B11-B20, D11-D20) and so on.

At portfolio review, tutors again work in pairs: Anna and Clara each review the other half of their students together (A11-A20 and C11-C20), Agnetha and Chrissie review their 'other half': A1-A10 and C1-C10 and so on.

Finally each two person jury compares and agrees marks. There is some difficulty in agreeing but finally they do it. Alternatively, they enjoy it so much they take even longer. Each tutor has seen 20 students at crit and another 20 at portfolio review. Each has had the chance to see the work of all their students and the entire output of another atelier.

Whatever they have agreed, a major problem is that each tutor has only seen the work of one other atelier besides their own. Consequently, two thirds of the year's output - and the approaches that produced it - are unknown to them. Subsequent year-wide moderation will be a slow, painful process.

Improvements

A few small changes can greatly improve both days.

How many cooks do you need?

The issue of how many staff are involved in assessment and moderation and how they work together is critical. There's a range of responses to this: each school tends to think their approach is the norm.

In some schools, every tutor sees every portfolio. With increases in cohort numbers that becomes increasingly difficult and time consuming. Elsewhere, limitations are agreed. Some schools insist work can only be assessed and moderated by tutors involved in teaching it: others resort to a final arbiter who is outside that group or have a 'star chamber' of particular staff.

Some schools insist a student's tutors are always involved at every stage: elsewhere it is argued that the student's work should stand up on its own so, for portfolio reviews at least, the work is reviewed by tutors who have not taught the student. Sometimes these rules are so embedded that they're not even recognised, let alone questioned or even explained to those new to the system.

Generally, fewer staff means quicker moderation. So, with the atelier system above, you could run two parallel moderating groups, each involving one tutor from each atelier. That way, work from across the entire year can be reviewed in half the time without any individual atelier feeling mistreated. Yet if half the year's teaching staff had seen a portfolio you could reasonably have some confidence in the assessment that resulted from it.

Rules & meanings

There is a tendency for schools to assume that the methods of assessment that they use are the norm. Nevertheless, visiting or part-time staff may not be familiar with these methods. Providing a clearer explanation of the school's methods can be a valuable and simple improvement to the moderation process. Explicit rules are simple enough to explain: implicit rules are sometimes just as important, but if there is no explanation given to new staff you cannot assume they are understood. Rules and timetable need to be explained and agreed in advance.

The good, the bad & the ugly

During assessment, certain portfolios become benchmarks, against which the other work is positioned. Typically these portfolios are at the high flyer, low pass, and high fail levels, plus benchmark portfolios at the borderline between classifications etc¹.

An important part of moderation is to get these benchmarks agreed by all staff. Assessment becomes easier and quicker once that has happened. That produces a circular problem: moderation can be done accurately and yet faster, by fewer staff or in smaller groups once benchmarks have been agreed: yet benchmarks are not agreed until everybody has seen all the work.

The trick lies in getting all tutors to agree the benchmark portfolios without this turning into the 'everyone sees - and argues - about everything' approach.

Three tips

The three tips that follow offer ways of ensuring that more staff see a wider range of work whilst still running to time, and of recognising, recording and addressing differences of opinion in a transparent manner.

1. The midday switch

Looking back at our earlier example, if Anna criticises Clara, Agnetha could criticise Duke. Between them, Anna and Agnetha would see all their own students, but would also see

¹ Note from editor: in some other educational fields this "norm-referenced" assessment technique is frowned upon. The latest theories suggest that assessment should be graded relative to a fixed set of criteria rather than by relationship to another student's work. Nevertheless, the complexity of design projects has made it difficult to create objective assessment criteria and so the creation of these benchmarks become a useful assessment tool.

samples of work from two other ateliers rather than one. Their marking would be done within a broader context.

If we switch the juries around half way through the day, tutors see an ever greater mix. For example, in the morning Anna and Clara see students A1-A5 and C1-C5. In the afternoon, Anna switches, and crits with Dave - seeing students A6-A10 and D5-D10. Clara also switches and works with Bjorn for the afternoon, seeing C6-C10 and B1-5.

This midday rotation has two notable side effects:

- the range of student work each tutor sees has doubled.
- timekeeping usually improves. Tutors tend not to finish so late. Perhaps because they are spending the afternoon with another colleague, tutors feel they must finish their morning's list on time. That means the afternoon starts closer to time, which means not such a late end to the day - causing less exhaustion and fewer frayed tempers and giving time for a tutors' meeting.

At the end of the day, if you make sure that each atelier tutor sees different ateliers' from those seen by their teaching partner then by the end of day's critting an atelier duo will between them have seen all their own students' work plus the work of four other ateliers. With our example, only the work of one of the six ateliers remains unknown to them.

Seeing a greater range of work in a shorter period of time builds moderation into the crit. Each tutor will prepare their assessment in the context of a wider range of work. Post-crit feedback given to students is in the context of having seen a more representative sample of the year's work.

That describes the process applied to a day of crits. If you apply the same process to interim crits and portfolio reviews, tutors will be making decisions and judgments on the basis of a much more diverse sample. At the cost of pre-planning and careful rotation of staff, much of the groundwork for speedy and accurate moderation has already been done.

2. The dynamic duo

The more people involved in looking at a portfolio, the longer it takes. If moderation means more people examining work, how can you run to time? One solution is multiple moderations by pairs of tutors, run in parallel, rather than reviews where everyone talks, marks and argues their way slowly through the year. Because this does not ask for - and does not want - extended discussion between tutors, assessment here can be much faster, and the pair of marks provides useful data.

3. Mind the gap

Once you have pairs of assessors or moderators working together, this technique makes marking quicker while also highlighting where arguments are likely and where there is a body of agreement. It can be applied to any situation where pairs of assessors or moderators are working together.

At the end of the crit, each tutor's sets of marks are collected. Percentages are needed to do the next step: if the school uses letters as grades then they have to be converted to percentages using a transparent, consistent method.

The informal comparing of grades or wheeler-dealing should be discouraged – this often takes place when tutors compare notes at the end of the day. That sounds counter intuitive since they are moderating when they do this. Tutors need to be told that it is better if they just record their personal assessment (this produces another time saving). The marks will end up reflecting tutors' views with a minimum of intervention from others.

Each student will have marks from two tutors. In the system outlined above, one will be their atelier tutor (the 'home' tutor): the other (the 'away' tutor) teaches a different group of students in the year.

The worksheet

Figure 1 shows an example of a spreadsheet that can be used for marking purposes. The first column is the student's atelier, the second column holds the anonymised students' names. Column three is the 'home' tutor marks, column four is the home tutor's name or initials. Column five contains the 'away' marks, column six is the average. Whilst the description of the worksheet sounds tedious, it contains no more than the information normally collected in less formal ways and saves considerable time later. An amendment of the document can also produce the day's timetable.

Atelier	Student	Home Tutor		Away Tutor		Average
		Name	Grade	Name	Grade	
Atelier E	Branson		41		43	42
Atelier F	Cayley		84		64	74
Atelier B	Durr		42		44	43
Atelier C	Eckener		76		72	74
Atelier B	Jong		60		50	55
Atelier D	Montgolfier		55		40	48
Atelier A	Nobile		32		52	42
Atelier C	Roe		56		54	55
Atelier A	Santos-Dumont		70		68	69

Figure 1

A worksheet records the mark given by each tutor for each student, which you can sort by student name, grade, atelier or tutors. Next, a column is introduced that averages the two tutors' grades. The process could stop here: the marks have been recorded and a final assessment produced by a fair and transparent method. Figure 1 is an extract from such a worksheet showing grades for nine students. As it stands, that table of marks does not reflect the disagreements between the tutors. The next step measures that.

Atelier	Student	Crit Home Tutor		Crit Away Tutor		Crit Average	Crit Difference
		Name	Grade	Name	Grade		
Atelier E	Branson		41		43	42	2
Atelier F	Cayley		84		64	74	20
Atelier B	Durr		42		44	43	2
Atelier C	Eckener		76		72	74	4
Atelier B	Jong		60		50	55	10
Atelier D	Montgolfier		55		40	48	15
Atelier A	Nobile		32		52	42	20
Atelier C	Roe		56		54	55	2
Atelier A	Santos-Dumont		70		68	69	2

Figure 2

One tutor's mark is simply subtracted from the other. If there is agreement the answer will be zero. Usually the answer will be a number, plus or minus. A formula is created using the absolute value (ABS) function to convert the answer to a positive number. For example, for students Branson and Cayley the answers 2 and 20 respectively would be returned. Bigger numbers mean greater disagreement between tutors. These figures are recorded in the Crit Difference column in Fig 2. Branson and Nobile have the same average mark, but while tutors agree about Branson, there is no meeting of minds about Nobile. A pass is 40%: one tutor regards him as a fail: the other as a clear pass.

Atelier	Student	Crit Home Tutor		Crit Away Tutor		Crit Average	Crit Difference	Folio Home Tutor		Folio Away Tutor		Portfolio Average	Portfolio Difference	Overall Average	Overall Difference
		Name	Grade	Name	Grade			Name	Grade	Name	Grade				
Atelier E	Branson		41		43	42	2		38		50	44	12	43	14
Atelier F	Cayley		84		64	74	20		80		60	70	20	72	40
Atelier B	Durr		42		44	43	2		40		44	42	4	43	6
Atelier C	Eckener		76		72	74	4		75		80	78	5	76	9
Atelier B	Jong		60		50	55	10		58		65	62	7	59	17
Atelier D	Montgolfier		55		40	48	15		50		40	45	10	47	25
Atelier A	Nobile		32		52	42	20		50		38	44	12	43	32
Atelier C	Roe		56		54	55	2		60		54	57	6	56	8
Atelier A	Santos-Dumont		70		68	69	2		73		65	69	8	69	10

Figure 3

If this technique is used at crit and portfolio reviews there will be grades from four tutors, a third of the total tutorial staff. All four grades can now be averaged and the Crit and Folio Differences added to give an Overall Difference column.

Grades for Eckener, Roe and Durr are remarkably consistent. Cayley, Nobile and Montgolfier must be looked at again. A threshold difference figure can be chosen, depending on time, and all work where disagreement is greater than that threshold is reviewed by another pair of tutors. Conversely, where there is agreement about work (Eckener, Roe and Durr) valuable time can be saved while having that work as a benchmark.

Other Ways of Seeing

The spreadsheet gives us other ways of looking at the grades.

Time & the midden

Assessment by pairs of tutors can save enough time to permit a quick, second round portfolio review using another pair, both 'away' tutors. Because the assessing teams are small and because they are not looking at their own students, results can come quickly. This approach can be used across the year's work - or just across folios where there is already high disagreement.

This third pair of marks means that work has been examined by six tutors - half the year's staff. Because of that, there is greater confidence attached to the averages. Consistently contrary opinions from all the remaining staff are unlikely: a lone wolf divergent opinion from one of them will not have a major effect.

See your world another way

Until now, the spreadsheet has been ordered alphabetically by student name. A copy needs to be saved so further manipulation does not run the risk of losing data: then time can be taken to order it differently.

Ordering it by Atelier and Total Difference may show that one atelier has higher differences associated with their students. This can indicate the subject matter or approach is 'out on a limb' without more moderation. The chances are that this has already become sensed intuitively. The advantage of this method is evidence - everyone else's figures - and this can pre-empt argument or ill will. Averaging and comparing the home and away tutor grades can also be useful.

Ordering data by Tutor and Total Difference helps identify tutors whose grades are consistently out of line with their colleagues. It also shows when individual marking behaviour is not consistent - and this is much more of a concern.

Advantages, Disadvantages & Queries

Strategic marking can be identified - tutor A is persuaded by tutor B to support student C so that tutor B will support one of A's students. The system of fast paired assessment and calculated differences does little in itself to prevent this, but it does make surreptitious negotiations more difficult. There is less time for horse trading and the marking profiles of individual tutors can be examined whilst also offering the chance for more assessments to be undertaken.

Applied across a cohort, reordering the spreadsheet and examining difference can help spot tutors who consistently mark high or low compared to their co-assessors.

More importantly, it highlights disagreement between tutors about individuals or the output of entire ateliers.

The system also shows where tutors agree. If there is high agreement, then there is little need for moderation and in a transparent way, students have been identified who need no further moderation.

Conclusion

This paper has suggested a number of techniques that can make assessment and moderation both faster and less traumatic. Some (the Midday Switch and Dynamic Duos) are time based and some are based on number crunching. Taken together, however, they suggest ways in which individual opinion retains its significance while contributing to a coherent view of a cohort's work.

There are of course more sophisticated statistical tools than calculating differences. This is true, but our method is easily understood and quickly applied. What it lacks statistically, it gains in simplicity and transparency - and that means it is more likely to be accepted and trusted by everyone.

These techniques do not - and should not - get rid of the essential occasions where everyone gathers around a portfolio to debate where it stands but it does help us identify where those events are most necessary and where time is best spent - and it saves enough time elsewhere to enable the job to be done properly.