**Formative Assessment & Standards Grading – Robert Marzano Workshop** at Hawker Brownlow Teaching & Learning Conference May 21, 2011

Please note: these are the notes taken during the presentation – presenter’s comments & ideas in black, my thoughts (Anne Doody), connections etc in blue.

Bob started the session by using the analogy of building a house – there is the start – plans, & end – when you get to move in, BUT there are lots of progress stages along the way that are evaluated (hopefully) for quality control. Eg. Slab, Frame, Brickwork, Plastering etc, Painting / tiling, Final Fit-out.

Thus, the summative assessment is when you get the key and move in, but if you don’t want ongoing problems, then the formative assessments need to happen along the way.

1968 – Bloom (taxonomy) tried to get people to take on the notion of formative and summative assessment in education – but no-one would listen.

At least this gives me some comfort – no wonder it takes years to get people to come on board and ‘act’ upon this consistently in their classrooms.

For teachers who are better at using quality assessments – it can move students from 50% to up to 84% in achievement.

He went on to emphasise that we should not lose sight of the importance of the interaction between teacher and student. If we focus too much on summative assessments – test results- then we (and students) can get ‘fixed on the number), and create ‘labels’ for students. We need to ‘use’ numbers, (data, Naplan, VCE etc), but also have a healthy ‘distrust’ of them.

Marks need to inform students of what they successfully demonstrated, AND what they yet need to know or understand.

**Forms of Assessment:**

* Obtrusive – tests
* Unobtrusive – teacher observation – sometimes students don’t even know they are being assessed
* Student – generated – when a student actually volunteers to show you – this can be really transformational

Bob suggests that it would be better if we could **‘assess more – and test less’.** He mentioned that Rick Stiggins has been talking about this for years. True – must check back to my own notes taken during a session with Rick Stiggins previously attended.

**Uses of assessment –**

Formative – can and should be used to track progress – he emphasized the importance of the tracking process for **both** students **and** teachers.

If we can SHOW progress over time – it is MOTIVATING.

Summative – the summative score for a reporting period should NOY necessarily be the last test in a time period – it can be an anomaly.

Instructional feedback is often not recorded, but is followed up on, and can made a big difference.

Marzano suggests – **we can’t rely on state tests or standardized tests** – then went on to explain why…

Cizek’s research:

‘Different scores’ when teachers look at subscales of whole tests and decide then what to work on. Like drilling down through Naplan to particular question responses.

Marzano suggests that it is **essential** to re-confirm judgements made from these analyses through **in-class** assessments **before** you act on using them to identify future foci.

Observed Score = true score + error.

Observed score – is the final score given.

True score – under ideal condition, where the student was calm, prepared, no distractions.

Error – this can inflate or deflate the overall score – it is when the student has either - become distracted during the testing for some reason; didn’t read the question correctly; OR guessed correctly, or the teacher rushed the correction process?

.45 is the typical reliability of a teacher – assessed test.

Educators live in an area of ‘soft science’ – psychology, sociology education etc. – we are trying to measure things ( learning) that aren’t easy to see.

At this point, we were asked to assign weightings to three different parts of an assessment so that the total would add up to 100. The three parts were:

1. Items 1-10

Ten items that require recall of important but simpler content that was explicitly taught.

1. Items 11 – 14

Four items that ask for application of complex content that was explicitly taught AND in situations similar to what was taught.

1. Items 15 – 16

Two items that ask for application in novel situations that go beyond what was explicitly taught.

After it became apparent that at every table individuals had chosen a variety of different approaches, he explained that:

All variations could be logical schemes, but that they show the importance of common understandings of assessment ratings for teachers teaching the same subject, especially to the same cohort / year level. It made obvious the differences between ‘teacher judgements’ that naturally occur unless we work collaboratively. This reinforced that research that Hill did many years ago, that showed that often the greatest differences were not between whole school environments, but rather between different classes in the same school.

This variation in interpretation, and as a result in assessments by teachers, can even occur when schools use rubrics – unless there is collaboration and thus a common language for devising them.

Marzano’s Rubric Proforma Style –

-this style enforces a process where teachers work through and identify ‘what is essential learning’ and articulate it clearly for students. Obvious link to McTighe’s ‘Understanding By Design’ work.

On a rating scale from 0 – 4, the target learning goal gets situated at Level/ Rating 3.

Regardless of the topic, the rubric has the following ratings:

|  |  |
| --- | --- |
| 4  Advanced learning goal | In addition to exhibiting level 3 performance, the student’s responses demonstrate in-depth inferences and applications that go beyond what was taught in class. |
| 3  Target learning goal | The student’s responses demonstrate no major errors or omissions regarding any of the information and/or processes (THAT WERE EXPLICITLY TAUGHT) |
| 2  Simpler Learning goal | The student’s responses indicate major errors or omissions regarding the more complex ideas and processes; however they do not indicate major errors or omissions relative to the simpler details and processes. |
| 1  Partial credit with help | The student provides responses that indicate a distinct lack of understanding of the knowledge. However, with help, the student demonstrates partial understanding of some of the knowledge. |
| 0  Even with help no success | The student provides little or no response. Even with help, the student does not exhibit a partial understanding of the knowledge. |

Marzano says that using these consistent ratings, and discussing in groups what each means in terms of the learning being delivered at any particular time, teachers can cut the time and complexity of rubric assessment by about two thirds over time.

This is because it is based on what becomes a common agreement of students’ responses being either:

TOTALLY CORRECT

TOTALLY INCORRECT or

PARTIALLY CORRECT

Thus building a methodology of true common assessment between individual teachers.

* Fascinating to be on a table with a teacher from NSW who indicated that there are no common rubrics used, no moderation during the teaching year, some, but minimal at the end-point – examination time. As a result, he felt that teachers have no idea how accurate their assessment of student work is against the judgements of others. While teachers in Victorian schools may frequently talk about how hard it is to find time to do this, at least the processes are encouraged, and in VCE enforced. Our progression points also assist with this.

Apparently Hattie has a similar thing / process that divides assessments into 4 quadrants – I need to check this\*\*

Marzano predicts that teachers will rely less on paper and pencil tests in the future – he assigned a rating of 34% reliability to these.

Projects – that may require a number of different rubrics to support separate parts – again – 34% reliability factor.

Probing discussions/ interviews/ conferences between teacher s and students – 67% reliable information gathered towards making a final assessment.

**Four Basic Approaches to building a final grade:**

Approach 1 – summative scores assigned at the end of the grading period. Can’t ever be a straight average of summative scores over time. Why? Early in the teaching period ( semester, year) students are likely ( expected?) to get low scores. This, in a way, is a period of pre-testing, because the teacher hasn’t done the teaching yet – thus early scores don’t can’t count.

Approach 2 – Gradual accumulation of a summative score. In this system, you stop assessing at a specific scale once you are reasonably sure that a student has attained that score value.

In my mind, this made me think of VET courses that focus on ‘Competent’ or ‘Not Yet Competent’. Once a student is assessed as ‘Competent ‘ they move on to new learning. This was reinforced when I conferred with someone else from my table who teaches VET in NSW – he explained that they have to keep reproducing assessments if students need them to in order to demonstrate competency. Other students keep moving on to new learning, while some stay on the same topic / learning to revisit content / practice until competency is achieved.

It also occurred to me at this time that ‘On Demand’ testing can be a quick guide for this.

Marzano pointed out that students’ starting point depends on their prior knowledge and skills, so it follows that students with lower starting points require more assessments along the way. This approach is labor intensive, BUT valuable.

For any given student, you use as many or as few assessments as are needed to make a valid and reliable judgement.

Approach 3 – The whole class progresses as one.

This approach was not explained in much detail – it seemed as though it was based on an idea where when students within a class was assessed, the top level possible was Level 2 of the Marzano rubric proforma. Remembering that this was:

Level 2 – The student’s responses indicate major errors or omissions regarding the more complex ideas and processes; however they do not indicate major errors or omissions relative to the simpler details and processes.

The inference from this was that the teacher thinks of achieving this level as success, and as soon as it is reached, moves all of the students on to the next learning. This may work well for organizing curriculum and planning – it does make one wonder about how this would fit with current focus on ‘student centred’ learning and teaching?

Approach 4 – continued improvement across the year of learning – with the additional provision of allowing students to go back and raise a previous score if they are willing to do so.

This approach seems to fit at least partially with the concept of Progression Points or Indicators of Progress on a continuum as we currently use – but – with tracking by the student to monitor self progress.

With this approach, the overall aim becomes for students to either:

Reach for Level 4 on the rubric – ‘*In addition to exhibiting level 3 performance, the student’s responses demonstrate in-depth inferences and applications that go beyond what was taught in class.’*

OR

A student can teach / tutor another student to reach Level 3 on the rubric – ‘*The student’s responses demonstrate no major errors or omissions regarding any of the information and/or processes (THAT WERE EXPLICITLY TAUGHT)’*

Robert pointed out at this point –

‘Eventually a [system] / district / school has to address the issue of report cards with some rigor and courage’. (This is the process that we hope is taking place at ACARA currently.)

I we then adopt the rubric system above, and take the step of equating it to the A-E scale that we persist with:

3.00 – 4.00 = A

2.50 – 2.99 = B

2.00 – 2.49 = C

1.50 – 1.99 = D

Below 1.50 = F (I’m so glad we don’t use ‘F’!)

To go even further:

3.50 – 4.00 = Advanced

2.50 – 3.49 = Proficient

1.50 – 2.49 = Basic

Below 1.50 = Below Basic

And for those who prefer % values:

4.0 = 100%

3.5 = 95%

3.0 = 90%

2.5 = 80%

2.0 = 70%

1.5 = 65%

1.0 = 60%

Below 1.0 = 50%

This certainly covers the many preferred approaches to arriving at final grades.

The suggestion to come next was that ‘Some braver districts and schools might wish to replace the time-based system with a performance-based system.’

Might this be what we would have had if everyone had had the confidence to truly rate each student against the progression points on every task – and we had not experienced the reality of the massive ‘C’ grade recording to see a much more diverse spread of student achievement?

Marzano was exploring the idea of passing students through who have not demonstrated knowledge / skills expected at a particular stage of learning. Eg. ‘D’ grade students still passed through to the next level. Obviously this opens the issue to other factors such as the social benefits of the same age (approx.) cohort – but did we really search for a way through this dilemma to allow students to really be learning at their own pace?

Robert said that ‘ when people get to the point in their career when they are capable of making the changes that they got into the career to do…they are feeling too worn out to do it!’

How do we encourage and support them to take the risk, now!

END OF SESSION

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_