

## Assessment as a gift, not a burden – Ben Lawless

You know, I am always really surprised when people turn up to hear me speak. Rabbiting on about assessment is not a popular conversation topic in many schools and I always clear the room when I bring it up at parties. Assessment gets a lot of bad press, for good reason. The purpose of assessment is murky. There's no shortage of top-down changes and ill-defined 'whole school focuses'. The high-stakes tests students do in their K-12 career bring anxiety and stress, and reduced Instagram. Most recently we have a focus on *data* and *spreadsheets* and *software* and other words that scare over 60s. Of course, I'm just joking. Even my nana can text now.... She actually sent me three blank texts just last week.

So, I was particularly stunned when the HTAV, the best teacher professional association in Victoria, if not the known universe, asked me to talk to you about this riveting topic. Deb Hull put it helpfully when she explained to me that a lot of what I say seems really obvious once you hear it. So, I will try and explain the obvious to you in a way that might make you see assessment in a new light.

Assessment has been a drain on student wellbeing and teacher workload for a long time. This is because for students the stakes are high and laden with judgement, and it hasn't helped teachers improve teaching. New understandings about assessment could change all of this. Now we just need to implement that research.

I'd first like to make some critiques about existing assessment practices, before I bring you all to the holy land of awesome assessment.

So... NAPLAN! What hasn't been said about NAPLAN? I think the debate around NAPLAN is an interesting case study because it highlights the general misunderstanding of the best use of evidence. How about this headline from the Age:

"State education minister will push for a dramatic overhaul of testing despite results showing Victorian primary schools leading the country."

Oh, come on James, don't stop us from being able to lord it over those other *dumb* states! We're better than other people! Surely that is the point of NAPLAN?

Two people I used to work with have made some interesting statements about it. Margaret Wu has pointed out that the margin of error for the test at the individual student level is larger than the expected gain for a student between tests. Translation: the test is practically useless for measuring one particular student's growth.

My old boss Patrick Griffin asked the question, if it's true that there's been no growth in the last ten years, why are we spending millions of dollars and wasting a week of school every year finding out what we already know?

Here's my question: at what point would NAPLAN achievement be sufficient for politicians and education-theory-illiterate commentators to stop tearing down our students, schools and teachers? Once every three-year-old knew the Theory of Relativity?

There are legitimate questions about the purpose of NAPLAN. If it is for information on how the Australian education system as a whole is tracking, why is every student in the country required to take it? People much cleverer than me have pointed out that we'd only need to test about 2% of students to get this information.

If it's for classroom use, why is the data not made available to teachers when it is produced? Now that we've got online NAPLAN (the one that ran as smoothly as the Ultrahet), surely we should get the data the next day? No, all the data was released when the pen and paper data came out.

NAPLAN data also can't determine teacher effectiveness, as Margaret Wu has also demonstrated

Publishing this data on the MySchool website compounds the problem, allowing people to make inappropriate comparisons between schools. The methods used to compare the schools are inadequate. The league tables that journalists produce by trawling through them is unsuitable, and the use of them by parents to decide which schools are better is mathematically incorrect.

But let's stop for a moment and imagine if NAPLAN *could* determine which schools were better. What then? How could we replicate the success of these schools elsewhere? As with many other issues in assessment, the problem is one of interpretation. How do we interpret the results of NAPLAN tests? How do we know what one school or one teacher did to get better results? Was it NAPLAN training? Was it strict discipline? Was it differentiated instruction? Was it an inquiry approach, direct instruction, free breakfasts, mood music in the classroom, expensive swimming pools, monogrammed boaters? Who knows? No-one thought to measure what schools actually DO to produce the results they get. Even when the Age recently trawled the MySchool website and found schools with high progress, they asked principals what they did and got answers like "discipline and explicit teaching". Hardly specific advice. One school said they didn't use a bell. How scientific! Now of course one could retort that we already know what works - evidence-based teaching strategies. But as Marzano eloquently states, not all strategies with high effect sizes work in every context. And it's a far cry from looking at a bland printout of effect sizes to interpreting them for your context, be that school, students, current teacher skill or whatever.

When I've discussed the idea that 'feedback' is an effective strategy, teachers all agree and reel off a thousand different things they do that they consider feedback. And it is not easy to disentangle the countervailing effects of different teacher actions, like teaching strategies which are more easily learnt, as opposed to teacher attitudes which aren't.

Another destructive fantasy is 'parental choice' in schools. The biggest in-school factor affecting student outcomes is teacher quality.

Between the three systems, state, independent and Catholic, there is hardly any difference in average teacher quality. But within each school there's a huge difference. Some research suggests that the best teacher in a school can be four times more effective than the worst.

Compare a student who gets the worst history teacher every year to another who gets the best. Or even more hilariously, the student who gets the worst history teacher every year at a school charging \$40,000 with another who always gets the best history teacher and pays no fees. Parents don't get to pick which teachers *within* the school their kids get taught by. So, yes, parents get choice, but it isn't about academic outcomes.

One of John Hattie's more cheeky ideas is to imagine what would happen if parents *were* allowed to pick which teacher taught their child within a school. Teachers who clearly don't enjoy the company of young people would be out of a job. A semi-related idea he had, was that, instead of performance pay, which hinders teaching quality because so much good teaching is collaborative and people would be fighting for limited bonuses rather than working together, was to take the autonomy of ineffective teachers away. Decide what a year's worth of growth looks like, pre-and post-test students, then those that get less than a year's worth of growth have to adopt the strategies of the best teachers.

Back to testing... All high-stakes tests have unintended consequences. NAPLAN is supposed to test the underlying skills of the country in numeracy and literacy. It was designed for system-level data. Parents getting their kids to complete NAPLAN practice tests completely screws with the data. NAPLAN is supposed to see

whether normal schooling improves numeracy and literacy. Practising for the test skews that data, because it isn't normal schooling.

If schools are ranked by NAPLAN results, and enrolments are influenced by it, it's completely rational for schools to prepare students for it. The first rule of any institution is to keep itself in existence.

I'm not suggesting do away with NAPLAN. But it should be transformed into a diagnostic on-demand test that teachers can run when and if they want, where they get the results immediately and hey – even some ideas for teaching interventions for students at different levels. Now I wrote that sentence before finding out about a plan by some federal agencies to do just that, but more on that later. For now, we don't need a national measure of attainment if the information is impossible to interpret.

Besides, we have signed up to PISA, which is more rigorous and less time-consuming. Our political leaders love PISA. Gotta be in the Top 5. It's like the Olympics. That warped institution that claims to be about fellowship but is really just crass patriotism.

So, what does it take to “win” at PISA? Singapore are currently winning at PISA. Do we want our school system to be more like Singapore? Let's see. Some Singapore high school kids start at 8am and finish at 6pm. How about the well-known statements summarising the approach to teaching and learning, “teaching is talking and learning is listening”, authority is “hierarchical and bureaucratic”, assessment is “summative”, knowledge is “factual and procedural,” and classroom talk is teacher-dominated.

Also, it pays to not include rural students in your statistics, who are socioeconomically disadvantaged and so perform worse. Be like Shanghai, Singapore and Hong Kong, where you literally have no rural or remote students. If Melbourne was a city-state we would no doubt be sitting pretty in the PISA rankings. Independence!

Many of the systems that do well in PISA do so because their students rote learn at the expense of arts, creativity and independence, spend hours in after-school cram sessions until after 9pm and have much higher class sizes. Even if these work, is our society willing to pay that price?

One group we regularly run into trouble with when trying to modernise assessment is parents. Parents, like everyone else in society, went to school, and so know everything about the education system and teaching. More importantly, they all went through a school system that used outdated assessment practices, like high-stakes end of year exams where you crammed content knowledge and regurgitated it onto a page. You know, like most university subjects.

Parents went to school where they got As for being intelligent and Ds for being thick. There was no way to go from one to the other. Getting an A was what you ‘deserved’. Growth mindset was not yet a buzzword. I can remember at high school, after noticing the girls were getting better marks, my male friends would simply say, “yeah, but they studied”. It's basically cheating.

There is a definite tension between old assessment practices and what we are moving towards. At parent teacher interviews, parents want to know ‘how their kid is doing’, code for, what rank do they have in the classroom. Comparing students to each other doesn't make them get better. Yes, some people are motivated by competition, but motivation alone will not get a student to develop higher order thinking skills, they need instruction or assistance.

In designing the best assessment system, I think we should be teaching parents what the most educationally beneficial things are, rather than letting their gut instincts guide our action. Instead of heeding the request to

lower class sizes, put all the brainy kids in one class, and give “deserving” students As, we should back our professional knowledge and explain to parents why we take certain actions.

We are unfortunate to be working in a profession where everyone thinks they know as well or better than we do. Again, because everyone *went* to school, they think they know how to *run* them. They don’t.

One of our *many* jobs as educators is to educate parents. We need to educate them about the purpose and correct use of assessment. We need to make them realise their experience of assessment at school as ‘judgement’ was wrong, and anti-learning.

In other industries, assessment information is simply that –information. Information to help find things out, information to make us ask ourselves questions, information to help improve things. In medicine and psychology, for example, assessment is used to make informed decisions. In education, assessment information is far too often used as judgement, specifically, negative judgement about schools, teachers and students.

So the main weaknesses of the understanding and use of assessment are firstly, we don’t know how to interpret the results of assessment, and secondly, we don’t use assessment information to improve teaching.

Assessment results can tell us a number of things, two of the most important being what a student’s current ability is, and how effective a period of teaching has been. The information about the quality of teaching is only relevant if we can work out *what about* that teaching was good, so we can improve the teaching of other teachers. Imagine two teachers, Shiv and Chloe. We pre-and post-test students. Shiv’s class only achieves a 0.2 effect size, but Chloe’s get 0.8. On this measure, Chloe is a more effective teacher. But why? The relationships she has with her students? Her sense of humour? Weekly pop-quizzes? Direct instruction style? Did she use learning goals and success criteria? Without explicitly recording what strategies different teachers used, and an honest appraisal of each teacher’s characteristics, like humour or empathy, we have no way of knowing. So we have no way of knowing how to help other teachers become more effective. So what was the point of gathering the data? Patrick Griffin always used to say, evidence not inference. Yes, we all have hunches about which teachers are good or not. But without evidence we don’t know for sure. And without evidence of their teaching strategies we don’t know *why*.

Most schools have dramatically increased the amount of evidence they gather and the time taken to do so. Teachers make the legitimate complaint that they’re gathering all this information and nothing is done with it.

But assessment should be for *teaching*. Assessment should diagnose a student’s current ability so we can teach them from that point. The best use of assessment is to target teaching and differentiate based on ability.

But does this happen? A teacher’s time is finite, and differentiating is extremely time consuming if you aren’t sharing the load. Our school system is obsessed with minimum standards, not equal growth for all students, so it makes sense for a teacher concerned about accountability to get everyone up to that minimum standard. So teachers teach to the middle and if they provide extra support, it is to get those bottom kids up to that standard. High achievers, well, “they’ll be fine”. Unfortunately this is kind of unethical. Every student should be making progress, and we aren’t really doing our jobs properly if we leave a group to their own devices. Research shows it is the top students in Australia that are making the least progress.

My experience suggests that most teachers teach the same thing to all students most of the time. I’m sure there is lots of innovative targeting and differentiating going on by great teachers in and beyond this room, but from what I’ve seen, this isn’t *most* teaching.

The recent experience of my own school is a good example. For years we had put modernising our assessment and reporting practice in the too-hard basket. Then about three years ago we bit the bullet and forged ahead, requiring all assignments to be assessed using rubrics written according to the guidelines developed at Melbourne University's Assessment Research Centre.

At first our collective skill in rubric writing was not great, but over time we improved. There have been some real gains: teacher assessment skill has developed, we have more student-outcome relevant discussions, our students are more literate in the meaning of skill development, and anecdotally they are developing higher order thinking skills. Our NAPLAN results have improved but everyone has their own theory about why that might be. Lots of people think it's the program they helped introduce. With so many factors at play, who can really say.

We've seen an increase in engagement, especially at the higher and lower end of the ability spectrum. High ability students have benefitted because for the first time we're explicitly telling them what it looks like to get better. Beforehand they could see they were getting A pluses and felt they were doing enough. Now the top criteria in our rubrics are 'stretch targets', things that are hard for even our best students.

Students with low ability are benefitting because they can see themselves moving upwards on a rubric, even if the marks they are getting are still Cs and Ds. They can see what they need to do to improve, rather than just how they compare to others, which is often demoralising.

This shows a huge benefit of a well-written rubric – it outlines the domain of knowledge. It shows students what increasing levels of quality in performing a skill actually look like. All the teacher effort of writing a good one with a team of teachers is worth it, as there is heaps of teaching *in* the rubric.

But again we come to the sticking point. Teachers gather all this data diagnosing student current ability but very few use it to target teaching.

I have spent the last four years slowly making a mega super-duper giant composite rubric with targeted activities designed to get students up to each level, which you can get for free on my website.

But even this is piecemeal. I only use this targeting strategy before each major piece of assessment. I do a quick pre-assessment, mark them against the rubric then they assign themselves two or three targeted activities to improve their skills. I do this twice per term if everything goes to plan. We're teachers, it hardly ever does. But in the ideal situation, the entirety of my teaching would be done like this. If I could convince others that differentiation and targeted teaching was a good teaching strategy, and get them to share planning, we could accomplish it.

We now use well-written rubrics and report on student work as it comes in, which some call continuous reporting or continuous feedback. We have stopped writing long wordy reports at the end of each semester. What's the point of writing stuff about a student up to six months afterwards?

As we all know, a lot of what gets written on semester reports is cut and pasted from comment banks and has little to no impact on student outcomes. Teacher time is arguably *the* most precious resource in a school, and report writing time is better spent elsewhere.

Teachers often complain about the time 'lost' to assessment, I interpret that to mean that they don't get to go off on fun tangents about what they personally find interesting in terms of content. I don't know about you but I find that much more learning occurs when we are working on assignments than when we are doing normal

classes working through content knowledge. I spend more time with individuals and small groups of similar ability. The students are more active and ask me more interesting questions about the content and the skills.

Historical skills including writing, analysis, observation and research are much more transferrable and useful than historical content knowledge, so spending time developing them makes sense. That might not be the wisest thing to say in a hotel full of history teachers, but I'm talking about the whole person here. Employers like students with liberal arts backgrounds because they develop thinking skills, not because they know that Cleopatra probably didn't really die from a snake bite.

There have been huge benefits flowing from this experience, but because we haven't yet gone to the next step of asking all teachers to actually use assessment evidence to change their teaching, we're missing out. Also, teachers make the legitimate complaint, "why are we gathering all this data"? Well yeah, why? If they aren't going to do anything with the evidence, if they aren't going to target teaching or make ability-based groups or mixed-ability groups with rich tasks and differentiated scaffolding, what is the point? They're just jumping through more administrative flaming hoops for what they perceive to be little gain. We need to win the battle for hearts and minds with all teachers. We need differentiation by ability to be in our DNA.

I once went to a rubbish PD on differentiation telling me that, hey you know, differentiation can be *anything*. I can differentiate with 'learning styles' or 'multiple intelligences' or by using colours or sounds or making different kids present differently. These might be great for engaging students and making class interesting, but they don't improve skill.

Differentiation must be based on student ability, giving students learning experiences targeted at their zone of proximal development, or 'goldilocks zone', where teaching interventions have the biggest impact.

Hattie distinguishes four levels of feedback – the product, a history essay, for example, the process, how to structure a conclusion, for instance, metacognition, like time management, and the self, like "good boy". My oral hygienist says "good boy" to me when she sees I've been flossing. It's nice. I like being a good boy. But it hasn't improved my flossing skills. Telling me how to floss, and doing a little flossing demonstration – not the Fortnite kind – would improve my flossing.

Feedback at the metacognitive level is best – teaching kids to think about their thinking.

Feedback at the level of the self has the least impact. And in fact, it has even more pernicious effects than this. Giving general praise is anti-growth mindset, making kids think they are 'good' or 'bad' at something in some static kind of way. It also turns them into praise addicts. They begin to do things to get praise, or grades rather than for the enjoyment of learning and to feed their own curiosity. It turns their motivation from intrinsic – for the love of learning, into extrinsic, for the desire to get told they are good, or to feel like they are better than others at something. Extrinsic motivation erodes intrinsic motivation, and when the source of the extrinsic motivation is removed – when the parent or teacher stops praising them – their desire to learn is lessened. Praising kids in a general way reduces the likelihood that they will be lifelong learners. My nutty mother had figured this one out when I was 11. I got a Mars bar from Mr Harris for doing well in class but Mum phoned the school complaining this would just make me want to do well to scam some chocolate, rather than for intellectual curiosity. *I never got another chocolate again.* But you know what, I am a lifelong learner. Thanks Mum.

There is a lot of talk about summative and formative assessment. There is an analogy that gets wheeled out to show you the difference. Imagine a restaurant that sells soup. When a customer tastes the soup, it is

summative assessment. They're assessing the yumminess of the soup and that's the end of it. But when a chef tastes the soup, it's formative assessment. She can adjust the flavour of the soup based on what she's tasted.

Except this is completely wrong. There is no such thing as formative and summative assessment.

It has been very confusing being part of the VCAA project titled "Formative assessment rubrics" when I don't believe you can divide assessment into formative or summative. It is like being part of the "Unicorn rubrics" project. They don't exist! But if anyone is recruiting, let me know. Assessment is not formative or summative, how you *interpret* the evidence is formative or summative. If I give my students a history test on the last day of the year and then say BYEEEE, put a grade on their report and that's the end of it, I've used that assessment summatively. If I give that same test at the *start* of the term, or *during* the term, and use the information to not teach certain things because everyone knows them, or focus on specific things because everyone struggles with them, or create ability-based or mixed-ability groups for some purpose, then I am using the test formatively. Using an assessment formatively helps you to *inform* your teaching.

To return to the soup analogy. If lots of customers taste the soup and say it's gross and way too salty, there is a good chance the chef might change the recipe. Likewise, if the chef is on her break and quickly downs a pot of soup, she's unlikely to bother changing the recipe after snacking on it. It's all about what you do with the information about the taste of the soup that shows if you are using it formatively or summatively. This point isn't just about the meanings of words. The difference between using assessment evidence formatively and summatively is huge.

We are all on a path of development. Hardly any of our assessments should be interpreted summatively. We've learnt from Carol Dweck's growth mindset paradigm that we should never see current skill, ability or knowledge as fixed. But summative interpretations do just that. ATARs can't be changed.

I don't deny that having a high-stakes end of year exam like VCE can motivate some students. But it is just as likely to cause stress and anxiety. VCE exams are, as far as I can tell, just a scam by the university system to rank Year 12 students so they can skim off the top 10% or so into the most prestigious courses.

The type of pedagogy required to prepare students for that assessment – a high proportion of teaching to the test, has negative impacts on the rest of secondary schooling. Earlier years are supposed to prepare students for VCE. So VCE curricula and focuses influence the middle years, Years 7-10.

The first time I ever gave a presentation at an HTAV conference was when I was showing a 'learning adventure' where year seven students learn about Ancient China by going back in a time machine and playing through a narrative adventure in teams of three characters, an inventor interested in maps and timelines; an artist interested in source analysis and a warrior-historian interested in everything else. I was stumped when a delegate asked me "But how is this preparing students for VCE?" How indeed. Either my completely amazing narrative history adventure was a silly teaching strategy or the VCE has a lot to answer for.

The ATAR is a poor predictor of success at university. And students with high ATARs, because they went to VCE-factory schools that spoon-feed students, do *worse* on average in first year university than those who studied more independently, but got lower ATARs. So ATARs become even more meaningless, which really is saying something.

So I hope we can all agree that learning is developmental. Therefore, it seems logical that our curriculum should be developmental – older students should have to learn more complex things. It is far from it. To be nice to Gerry Martin, our wonderful History Curriculum Manager for Victoria, I will critique the Australian Curriculum instead of any state curriculum, but this point is still valid for all state curricula as well.

Some easy to find examples show what I mean.

Here's the Year 5-10 whole skill curriculum in a table, the way you actually want to see it rather than a 38-page document.

Here's my favourite, "use historical terms and concepts". A student in Year 5 is only capable of 'using historical term and concepts', whereas the mature, VCE-ready 16-year-old can, well, 'use historical terms and concepts'. Okay sweet, that shouldn't be too hard to teach.

How about research skills? Our naïve 10 year old youngsters have to be able to "identify and locate a range of relevant sources". Sounds tricky! Six years later... do that but *also* "use ICT and other methods". Great. I wonder how they were locating those sources *before* they had to "use ICT or any other method"?

We cannot rely on the authorities to produce developmental curriculum, at least not in a timeframe that helps our existing students. We need to do it ourselves.

VCAA do note that you don't have to teach the Year 7 curriculum to Year 7 students, they say you should teach them at the appropriate level. That's problematic. Firstly, you aren't going to teach different *content* to different kids in the class. I'm not going to have some kids working on Ancient Greece while others learn about Medieval Europe. Secondly, the curriculum isn't developmental anyway. So if the departmental advice is to get students to learn from a higher year level curriculum, there is no guarantee it is more complex.

At our school, we use the curriculum to determine which skills we teach, but ignore the age-based expectations. As you well know, there is an enormous range of ability within one class, sometimes up to eight years of difference. The curriculum's published age-based norms are usually relevant only to about the middle 30%.

We write assessment rubrics that have the potential to diagnose student learning by making the lowest criteria something everything can do, and the highest criteria something that is a stretch even for the most able student. Anecdotally this approach has had success, as many of our teams have needed to increase the difficulty of the highest criteria, meaning our top students are getting better at higher-order skills.

There have been some unexpected side benefits from this approach. The quality of the discussions between teachers at our school has gone up considerably. We now discuss how we interpret the curriculum, what skills and knowledge we think is most important, and what student development looks like.

Our understanding of evidence use has increased, and the rubrics we write are actually useful at diagnosing student point of readiness rather than the atrocious ones that can be found almost everywhere.

We have started on the journey of reporting on progress, not achievement. This is something Melbourne University's John Hattie and ACER's Geoff Masters think is crucial. There have in fact been motions put in place by a group comprised of the leaders of ESA, AITSL and ACARA to develop online adaptive tests that would place students on a developmental framework. The "National Learning Progressions & Online Formative Assessment Initiative", or NLPOFAI for short.

This is great news. For an assessment geek, this is about as good as it gets. But this is only for numeracy and literacy. And, the 'discovery' phase of the project alone is going to take six months and the business case document came out almost three years ago. So, don't hold your breath.

In the meantime, using just excel spreadsheets, we have been able to produce our own developmental progressions of skill which span across years, backed up by data. Soon we will be able to report on student

progress up these levels, stating what level a student is at, accompanied by a simple one-sentence level statement appearing on reports. Way better than giving out an A-E grade. Everyone will be able to see actual progress in skill level across year levels, this one is from Year 7 to Year 10.

There are a few steps to producing a framework like this but it isn't rocket science. Get in touch if you'd like to be able to generate this kind of thing yourself.

So what would my ideal scenario look like, using what I consider best practice assessment?

The year would start off with a pre-assessment per subject. These would preferably be automatically scored and linked to learning activities. After the pre-test, a teacher would get an easily digestible report on each student, on the class as a whole, and on potential ability-based groupings in their classes. These reports would place students on a developmental progression of skill for the subject area. With the reports would come links to targeted instructional activities appropriate to develop student skill. These could either be self-regulated, small group or whole class activities. Course content material would also be available at at least three levels of difficulty. For part of their educational experience, students would work through modules of activities to develop their skill and be exposed to curriculum content at a level appropriate for their current literacy. The activities would either self-grade or be quickly graded by teachers.

If all this material was planned collaboratively, I think activities at different difficulties and content knowledge material written at various literacy levels would not be that hard to make.

Assessment would be more frequent, quick, small and low-stakes. Failure would just help pinpoint where teaching is most required, not to catch a student out and result in a nasty letter grade on their 'permanent record'.

I personally don't feel the large piece of work students might do, an essay, for example, should be a source of summative grading by the teacher. The fear of failure in high-stakes assessments like these make students play it safe, learn less, take less risks, and generally suck up to the teacher, rather than embark on their own intellectual adventure.

I would still like to see an evaluation stage though, but this could be done by the student themselves, promoting metacognition, or from other students. Peer learning helps, as peers give fair but generally positive feedback that other students understand, whereas a lot of teacher feedback is overly abstract and critical. If you've provided feedback and direction along the way, I personally don't see the need to judge their output. Let *them* ascribe value to it.

In the clinical teaching model, borrowing from the medical fraternity, assessment is used as information, not judgement. The clinical model of teaching, modelled on clinical medical practice, outlines a process like this:

We use assessment to find out what the learner is ready to learn. How should we teach them from this point? Pick a strategy that should develop the student's skill. Think what should happen if the teaching was successful and how you can measure that. Teach the student. Did it work? Awesome, repeat the cycle.

Of course, what I am suggesting would not be the only thing that took place in the classroom. We teachers do a million things other than share knowledge and develop skill. We motivate, inspire, amaze, create lasting memories, develop character, engage, generate curiosity and introduce new ideas, while hopefully leaving time for calm reflection and the birth of wisdom. All before recess.

As the semester or year progresses, teachers record the teaching strategies they used. Perhaps using some kind of electronic form with tick boxes, so as not to increase workloads.

At the end of a teaching period, students would sit a post-test. These results would be used to assess teachers, teaching strategies, courses and individuals.

We teachers should be secure enough to put ourselves through the same assessment and improvement cycle as our students. We need to move away from a culture of simply sharing teaching ideas to a culture of challenging teacher ideas. If we can show, in a vaguely objective way, that some teachers are more effective, we question the teaching behaviours of the less effective. I definitely think this has been a barrier to data use in schools - the idea that data might uncover ineffective teaching, when it has always been so easy to hide by simply closing the classroom door.

Ask which teachers were the most effective? Why? Is it the teaching strategies they used, something about their personality, something else? Is it transferrable to other teachers? If so, spread it.

Ask which strategies that multiple teachers tried were most successful? Get more teachers to do that.

Which courses produced the highest growth in certain areas? Interpret why this might be. Maybe your Year 9s improve greatly at chronology. What is it in that course that helps them improve? Do that elsewhere.

For individuals, teachers could question what they did that worked well with certain students, or why other students didn't show that much growth. Did they have outside of school issues? Was the teaching style not suited to that student? What can teachers learn about their own teaching from their students' results?

The Education Act requires schools to report twice a year against a minimum five-point scale. I would report student progress to parents using a rocket report, with simplified level statements taken from the underlying developmental scale.

This would show what a student can do and what they need to do to get better, not give them a letter or number that is impossible to interpret in any meaningful way except to say, "you're hopeless", or "you're great, don't worry about pushing yourself".

I think the best use of assessment is as *one* tool in the arsenal of dynamic and responsive teachers, so they can diagnose student ability and teach them in the goldilocks zone – not too hard, not too easy. So many of the other uses for assessment are to achieve other goals, which are nowhere near as closely connected to learning. Let me end by telling you how I think we should use assessment.

Only gather data if you can interpret it, by also recording what teaching strategies and other teacher behaviours you used to get those results. Act on data that you gather by changing your teaching – target instruction, make ability or mixed-ability groups. Start making and using developmental progressions. Report on student progress rather than just current achievement levels. Start having discussions with people who have outdated views on assessment whether they are a parent or your principal. Convince them that assessment can be a gift, not a burden.