Denis Fitzgerald

DENIS FITZGERALD suggests a new path forward for assessing students in our Public Schools…

Where are we?

This is not an Apocalypse. Nor an occasion for melancholy. It might, however, be a crossroads.

Too many teachers and too many parents are now reporting that schooling is changing – narrowing, tightening, worsening. And this is across schools and all systems. It appears that the testing and data bombora has possibly reached a critical mass and has begun to somewhat corrode the life and richness that the best of schooling has provided to children for generations.

There was a time when people from education departments visited public schools, scrutinised what was occurring, suggested how improvements might be made, promoted those considered engaging in exemplary work and cross-fertilised good works and ideas between schools and localities. Actual people. Expert in education.

Those times have gone. And with 'efficiency dividends" and populist calls to scale back bureaucracies and "back-office staff" it is not likely that those times will re-appear soon.

In their place we have the "science of management", the principles of accountancy and the litany of "evidence-based" approaches and the glory of data. Mountains of it.

Naturally, data, in the form of numbers, has a rationalist appeal. It is readily collected, transported, stored, and quite seamlessly can have meaning and significance ascribed to it.

Departments and Governments seem to be developing education policy with a zealous emphasis on data and numbers which is setting directions contrary to the long-term interests of students.

Under the rubric of forming policy that is "evidence-based", these fashions are reducing the complex and human dimensions of educational thinking to a dull ciphering that robs our public system of its focus on inculcating the highest of skills and the finest of values amongst our schoolchildren.

This tunnel vision is increasingly being seen in a broadening call for more testing and yet more test data and the constant reporting of such to the centre and the media. Schoolchildren are being tested more and more in nationwide assessment regimes, whilst state-wide systems of support for students in need are in rapid decline.

Denis Fitzgerald

To reduce policy formulation to a mantra about "evidence", which has become a pale synonym for tables and integers diminishes all who are involved in public education.

Sound education policy does indeed emanate from evidence, but the nature of such evidence is diverse, complex and pluralist. It does rely in part on assessment and data, but it is much more.

Show us the evidence

Evidence in education, as Dr Paul Brock outlined, blends teachers' capacity for observation, classroom and human experience, the focus on children as individuals, a reliance on authentic research which is not partisan or tendentious, a commitment to genuine scholarship, the deployment of acquired wisdom and the practical, daily application of nous.

Evidence in education needs to concentrate on cause and effect; it appreciates context; it avoids the singular magic bullet; it disregards irrelevant overseas policy recipes; it understands complexity and it avoids sloganeering and simplification.

Appropriate professionally-directed assessment, testing, reporting and feedback are all elements of sound teaching practice but they are properly used as cornerstones for intellectual confidence and growth, creativity, engagement, innovation, imagination and speculation.

The current dull empiricism of collecting neat piles of test score numbers often restricts teaching to mindless arithmetic, where classes become databases. It distracts from the expertise-based reflection and experience of teachers and principals and fails to recognise or value the practical, strategic commonsense and awareness that underscores good classroom practice and intelligent policy formulation.

The emerging fad of target-setting, based on using mass diagnostic test scores, commits the fundamental error of failing to understand the purpose of true diagnostic testing. When a measure becomes a target, it ceases to be a good measure. Its purpose has been corrupted and invites the gaming of assessment systems that ought to be solely focused on identifying and then remedying the intellectual needs of children.

We have moved beyond the stage of the proliferation of tutoring and heightened stress levels for students beginning in the primary school years. We now stand uniquely in the world with our test obsession, where it is now even commonplace for our supermarkets to do thriving business in NAPLAN test preparation publications for our very young. Sometimes to be found in Aisle 3, just next to the medications.

So, given this, how might we approach our crossroads?

Denis Fitzgerald

Some assumptions

This paper starts with the view that sound assessment practices are vital to a quality education. It does not embrace the progressivist view that the processes of assessment and testing are essentially oppressive to children. In fact, constant assessment is an intrinsic element of good teaching practice. Good teachers are in the process of assessment from the time they enter the school car park in the morning.

Testing, similarly, has its place in order to determine how children and communities can be supported, their particular needs identified and resources dedicated to their educational care. For disadvantaged students and communities, the processes of assessment are particularly important in building resilience and as a starting point in focusing on need.

The essence of the problem has actually been centred on how assessment and testing outcomes have been reported. In national testing regimes; in the media; in ill-informed political commentary; at yuppie dinner parties.

As Professor Roger Murphy of Nottingham University observes of his country:

> Many of us would argue that there is nothing intrinsically wrong with occasional classroom-based tests or even, for certain purposes, national testing programs. However, many would also agree that such tests have limited value, can give a misleading and partial view of educational progress, and, if the scores are given too much value and importance, they can lead to dangerously distorted teaching and learning and seriously poor judgements about pupils, teachers, schools and localities."

And as England continues to be the fountainhead of dull educational policy, we heed also the words of Kevin Courtney, who reflects on where the testing-obsession has taken that country. Assessment mania there, according to Courtney:

> Has become a notorious example of teaching to the test: it narrows the curriculum; it prioritises the production of test scores above the support for children's learning. In some cases it damages children's sense of well-being. The burdens it imposes on teachers are unjustifiably heavy. The aspirations of teachers and the capacities of pupils are frustrated by a system that is not fit for purpose.

The data messiah

Across the Atlantic, we see an analysis emerge that bears striking similarities to what England has endured. According to Shepard, Penuel and Davidson in a recent edition of *Phi Delta Kappan*:

Denis Fitzgerald

> The testing mandates of the past two decades have only made things less coherent (or coherent but not meaningful, insofar as they have fostered a teaching-to-the-test approach, aiming towards narrow curricular goals).

These authors are especially relevant to Australia with their critique of data-driven decision making. Their research leads them to conclude:

> Data-driven decision-making is most accurately portrayed as a policy *theory of action* [their emphasis]. It relies on no specific model of learning but, rather, draws its inspiration from theories of organizational change …

> Data-driven decision-making assumes that teachers will know how to help students – or will seek training that shows them what to do – if the interim tests reveal that students are struggling. But this assumption has never been supported by empirical research findings ... and researchers have found this approach to be especially ineffective in low-performing schools that tend to lack the capacity to adapt in this way (Elmore, 2003).

> To date, most of the research on data-driven decision making has focused on the work of data teams (groups of educators tasked with analysing test results). Findings show that, at best, such teams are able to identify which students are most in need of help and which objectives are most in need of re-teaching...

> However, because interim assessments offer little or no insight into the reasons why students are underperforming or how to help them, their use hasn't been found to lead to improvements in teaching or learning … When students are struggling, being told how far behind they are does not help them move ahead … Instead, students need meaningful opportunities to engage with the material, ask questions, try ideas and receive useful guidance and feedback from teachers and peers.

What these distinguished US researchers conclude in their work on design principles for new systems of assessment is clear, and for we Antipodeans, telling:

> To be valid, reliable and affordable, state accountability tests must be standardised, and this often makes them ill-suited to serve as models for high-quality teaching and learning at the local level.

> A wiser approach, we believe, would be to start *not* [their emphasis] with statewide accountability tests as the primary driver of educational reform but to begin, instead, with local decisions about curriculum and instructional practices, informed by small-scale (and usually low stakes) assessments that are grounded in a single, coherent model of learning that is consistent with contemporary research findings about cognition, child development, motivation, identity formation, and equity-focussed instruction.

This complements the understanding of Dr Paul Brock in his thesis of what constitutes real evidence in education – genuine scholarship, authentic research, acquired wisdom and nous.

Denis Fitzgerald

And in terms of these four criteria, we might be wise ourselves to dwell on what Dylan Wiliam understands as the role and scope of assessment. Put simply, assessment is the bridge between teaching and learning. For Wiliam, assessment is the measure of our effectiveness, our professional account delivered to those who provide funds and an indicator on what our students retain.

Given its importance though, we as educators have a responsibility to honestly proclaim when assessment systems are in error, are misguided and are causing damage.

Dylan Wiliam is very much of that school. He makes it plain:

> In every single instance in which high-stakes accountability systems have been implemented, adverse unintended consequences have significantly reduced, and in many cases have completely negated, the positive benefits of such an assessment system.

Wiliam recognises how the longer centralised test regimes are in place, the more certain they are to be gamed. Teachers become better at narrowing down the scope of their teaching and that which is taught and tested becomes more specific and predictable. But of course "the performance of the students on things that were not tested got worse." And we know how often this has led to the neglect of the creative, physical, imaginative and performing arts -- the wellsprings of humanism, the intangible and the unquantifiable.

Instead of preparing students for a range of life choices – including diverse work and career trajectories – the narrowing of assessment restricts what is taught and valued in our classrooms and constricts also the scope of an individual's human possibility over time.

There would scarcely be a teacher in Australia unaware of this identical process shrouding schooling in this country.

Other pathways

If we return to the image of the teacher entering the school grounds at the start of the day, we can remind ourselves of the constant reality -- that teachers are always assessing. They reflect and observe constantly -- who has arrived at school; what that cluster of students might be contemplating; the vacant pupil desk in the classroom; the look on a child's face in the corridor; the telltale linguistic gem inside a piece of writing; the anxious parent at the front office; the child uncharacteristically unkempt one morning …

Subconsciously, teachers are assessing all the time; forever gleaning "evidence" from a range of sources. Such gleaning reflects the broader understanding of what the school curriculum is for a child – the full set of human, emotional, ethical and intellectual interactions the student encounters within the school

Denis Fitzgerald

day. Teachers are forever exercising their professional scrutiny and judgement upon these processes.

To begin constructing a replacement for the current failed assessment and testing mess, we have then a starting point for our understanding of the diverse and complex nature of the broad curriculum that students experience and the different contexts in which teachers operate.

As Wiliam observes of the more formal assessment processes:

> The challenge, therefore, is to create an assessment that is externally-referenced, distributed and cumulative. The assessment system needs to be *externally-referenced* [his emphasis], so that the teacher can honestly say to the student: "These are not my standards." When the authority for the standard of achievement that students have to reach does not come from the teacher, the teacher is free to be a coach, rather than judge and jury.

> When the assessment is distributed across the whole course, the negative effects of "teaching to the test" are minimised, because the only way to maximize the students' achievement is to teach all the students everything. When the assessment is cumulative, there is no incentive for students (and teachers) to adopt a shallow approach, because if material is forgotten, it has to be learned again, because it is going to be assessed again …

> The important thing is that the assessment system, as far as possible, creates positive incentives for teachers to teach well, and for students to study well. Once this kind of assessment is in place, it should fade into the background and be unnoticeable and unremarkable.

In the Australian context, this would all mean that the uniform mass testing of de-contextualised skills and knowledge would need to be replaced with testing that has a range of intellectual demands for different student capacities at different times and would be completely embedded in syllabuses and curriculum developed by the appropriate statutory bodies. In simple NSW form, it would require assessment systems to be differentiated, to be derived from what exists in NESA syllabus documents and for teaching programs to be solely derived from these syllabuses. Quality, equity, personalised learning and curriculum-assessment alignment would all be enhanced through this process.

Such an approach would allow schools to progress beyond the current modish managerialist obsession with number harvesting, spreadsheets, datawalls and scattergrams. As Wiliam's research establishes: "There is absolutely no evidence that this kind of monitoring has any impact on students' learning".

What it does do however is give the illusion that something constructive is being done, that a pseudo-scientific method is being applied to student need, that political accountability is being served and that the human mind and its progress can be simply mapped. It is scarcely the much-heralded twenty-first century approach to learning. In fact, it bears the hallmarks of a nineteenth century approach to schooling, rather akin to phrenology.

Denis Fitzgerald

Pasi Sahlberg and Jonathan Hasak remind us that this latest datawall fad is yet another import from US schools where they adorn corridors and staffrooms. They conclude:

> These data sets, however, often don't spark insight about teaching and learning in classrooms; they are based on analytics and statistics, not on emotions and relationships that drive learning in schools. They also report outputs and outcomes, not the impacts of learning on the lives and minds of learners.

World-class nous

The good news is that Australia has an especially strong research base in outlining what does indeed lift student achievement in all school communities, especially those in greatest need. An exemplary case of this is the work of Professors Wayne Sawyer and Geoff Munns and their colleagues in The Fair Go Project that has closely analysed the teaching practice of highly successful schools in NSW.

This project has identified the consistent themes and characteristics of successful teachers and teaching. It found that what truly works is a whole school approach that has a consistently high level of intellectual challenge and where actual teaching and learning are central to all classroom activities thereby forming the basis for academic success. This is more elaborate than that simple exposition suggests however.

The successful classrooms are those, as Geoff Munns describes them, where: "The schedule is demanding, with expectations that all learners will be involved with work that is cognitively challenging. The combinations of intellectual quality, active and authentic learning and consciously designed enjoyable experiences build engagement".

Flowing from this, the classroom becomes a place of:

- Higher order thinking, problem solving and analysis;
- Research and experimentation where students are encouraged to question their conclusions;
- Student-centred discussion;
- Teacher modelling;
- The foregrounding of goals, concepts and criteria;
- Linking existing student knowledge and new fields of enquiry;
- The integrated use of resources, including ICT, with rich tasks to encounter;
- Relentless questioning;
- The embrace of intellectual risk;
- Valuing the process of learning as well as content knowledge;

Denis Fitzgerald

- Building understanding and answering problems;

- Interpretation rather than reproduction of rote knowledge.

The Fair Go Project concentrated its work in low SES communities and was able to demonstrate the particular efficacy of these approaches in schools where teaching and learning has not always flowed trippingly.

Encouragingly, schools that have followed these positive pathways of thought are demonstrating clear success. Woonona High School, as one example, has a whole school approach of rich task design, high standards in all classrooms, a practice of drawing students into their own assessment and a culture of setting and achieving challenging learning goals. And it works.

It is of particular significance that what has not benefited low SES school communities has been the testing agenda of the last 25 years. Indeed, these communities are some of the biggest losers, globally, from the test obsession.

As Linda Darling-Hammond reports from the USA:

> Poor districts offer stripped-down drill and practice approaches to reading and math learning, rather than teaching for higher order applications … Critical thinking and problem-solving, collaboration, effective oral and written communication, accessing and analyzing information, curiosity and imagination … have typically been rationed to the most advantaged students in the United States.

What Darling-Hammond identifies here as higher order needs are the very skills that youngsters from linguistically struggling home environments need to have explicitly taught, valued and practiced in their classrooms.

Alan Luke describes a similar phenomenon in Queensland: "Far from students in poorer communities lacking 'basic skills', they in fact receive more work on decoding at the expense of other critical aspects of reading and literacy."

Teese and Lamb draw this conclusion from their national research: "Since public perception of schools based on league tables particularly disadvantages low SES schools, the consequence is a focus on 'performance', rather than 'achievement'."

Meanwhile, both performance and achievement are in decline. John Hattie observes that PISA results have steadily declined since the year 2000 and this has been principally because of the worsening relative outcomes of the top 40 per cent of students.

And here we draw closer to the quintessential failures of the testing regime. It does not focus on deep performance. It does not enhance performance. It has corresponded to an inexorable decline in results across the board for Australian students

Denis Fitzgerald

in international measures. It correlates with a cross-curriculum decline in top end outcomes, without any commensurate lift in bottom end achievement. It has narrowed the taught curriculum. It has sapped the joy from too many classrooms and programs. In fact, it has conspicuously and remorselessly failed by every criterion and goal that it has set out for itself.

Perversely, our sustained national decline across stages, years, subjects and learning areas has led some key policy makers and politicians to prescribe the need for more testing as the cure for the very malady that a generation of testing has clearly contributed to.

One might suggest therefore, that the wave of testing has nothing to do with direct student benefit or enhancing student achievement. It now borders on ideology.

Naturally, it is important for schools and systems to review the outcomes of the students in their care. Indeed, systems have an ethical obligation to use such data to apportion resources as they would in any other area of public policy in direct ratio to areas of identified need.

This is another question entirely, however, from actually using this data for the purposes of bloodsport media reporting or mindless yuppie chatter. To use the private academic performance of children in obligatory tests as the basis for public speculation, social division, community shaming and talkback prattle is shameful. Systems can collect all of the data required to inform education and resource policy, without reducing the process to the public spectacle it has become today.

What really works

Any responsible policymaker is aware of what works best to enhance student achievement. John Hattie's empirical work, for one, lists these as the top six effect sizes:

1. Formative Evaluation
2. Teacher Clarity
3. Reciprocal Teaching
4. Feedback
5. Teacher-student relationships
6. Spaced versus Massed Practice

Each of these requires elaboration, but the directions and essential directions are clear.

The NSW Department's Centre for Education Statistics and Evaluation [CESE] in its 2014 publication:

Denis Fitzgerald

What Works Best drew similar conclusions and identified "Seven themes from the growing bank of evidence we have for what works." Their themes were:

1. High Expectations
2. Explicit Teaching
3. Effective Feedback
4. Use of data to inform practice
5. Classroom Management
6. Wellbeing
7. Collaboration

Taken together, these compilations of high order research are notable for their range, diversity and the almost complete absence of mention of high stakes testing and assessment as a means of enhancing student learning and achievement. The use of data is given modest prominence and only in the context of informing teacher practice and reflection as an adjunct to informing professional judgement.

With similar enlightenment, NESA syllabuses in NSW also manifest profound Assessment Principles from a student-centred, higher plane. They seek to guide teachers to an understanding of assessment in its three roles: as Assessment For Learning, Assessment As Learning and Assessment Of Learning. Sadly, it is the last leg of this trifecta that attracts so much of the attention.

To return then to Paul Brock's four points about what should constitute an evidence base for the finest of teacher professional performance we see in these findings and approaches of CESE, Hattie and NESA an affirmation of his blend of genuine scholarship and authentic research, along with the application of acquired wisdom and nous.

Regrettably, this guidance from NESA, Hattie and CESE is now more notable in its breach rather than its observance. What is occurring in too many classrooms is dull test preparation, based rarely in the spirit of syllabus documents and best practice. What germinates too often is rote, low-level test preparation as schools gird themselves for the autumn of testing and the winter of reporting.

The provenance of error

Then why has it all come to this? Of course, much has been written about the unsuitable but ubiquitous application of capitalist managerialism in education policy settings and practice. It is clearly not fit for purpose. It misunderstands the nature of the relationships between teachers and students and schools and communities. We have had a generation of that managerialism in Australian education. And it has failed. But I will leave that deep critique to reside elsewhere.

Denis Fitzgerald

The mislocated application of management theory to assessment and testing has other intellectual and ethical flaws. It fails by its own criterion of deploying an inappropriate scientific method.

Dr Alpesh Maisuria of the University of East London explains this fatal flaw thus:

> The philosophy behind testing in schools is a false application of approaches used in some of the natural sciences. In the natural world, through experimentation, we can observe and uncover the fixed laws of nature. This has allowed scientists to predict with a high level of accuracy the outcome and regularity of what happens in the natural world, for example the combination of two hydrogen atoms with an oxygen atom will always result in water being created.

> This level of certainty and predictability does not occur in open systems such as weather and climate, nor is it appropriate when describing and explaining children's learning in schools.

> A positivistic logic has been misappropriated and thus we mistakenly expect standard and predictable responses in tests from humans. The idea is that we provide a standardized education for all and expect comparable outcomes.

As anyone who has spent meaningful time in a classroom knows, they are certainly not fixed systems and the youngsters there are complex human beings who can scarcely be described or captured within a test result.

As Maisuria puts it: "Tests only give a surface metric, rather than deeper understanding of what has been learnt, why learning has taken place and how. In short, we have a misplaced trust in the accuracy of data from standardised tests."

What is especially concerning however, is that the centrality of testing is not only deepening but its longevity means that there is a generation of teachers, across all schools and systems, who have known no other central narrative in schooling. They learnt about it at university and have been immersed in its endurance and ubiquity from their first day of teaching.

Teachers and principals and school communities have also seen, over this time, the grave damage that thoughtless system and media reporting of these test results has caused to vulnerable communities and students. Consequently, and with the best of protective sentiment, schools without any particular affection or trust in the testing regimes have thrown their lot into seeking to ensure that their students are not the ones to be this season's victims. And so they teach to the test. They have special sessions and classes and tutorials and experts and targets and strategies and staff meetings and practice tests to get their kids over the line. Sometimes they come up green and sometimes they come up red.

Additionally, well-intended teachers might occasionally determine how they might game the system – to seek to predict the questions or focus areas or text type or detect patterns of emphasis and generally

explore the tea leaves of past assessments to give their kids and school an edge.

So how did it all come to this?

Denis Fitzgerald

Professor Richard Pring of Oxford University sees it all as an inevitable end product of the managerial approach to education and our new meta-language of targets and delivery and performance indicators and audits. He understands this dire present as being pre-determined: "The inevitable problems associated with undue weight and emphasis on a single indicator for monitoring complex social phenomena."

Pring cites two "Laws" derived from public management principles and economics to further assist in the explanation.

He cites Campbell's Law, which stipulates: "The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it was intended to monitor."

And he adds Goodhart's Law: "When a measure becomes a target, it ceases to be a good measure."

Even by the principles of management theory, the testing regime and attendant culture has its own inbuilt and inevitable failure and the longer it endures the deeper is its failure and the damage it may cause.

Central to this damage is how it corrodes the nature of the classroom, the relationships within it and the purposes of schooling. As we are directed to delve more deeply into the data obsession and integers as evidence, the more we see teaching and learning divert from teaching the whole child the whole curriculum.

Under this dispensation, classrooms become data farms and principals are encouraged to become more an unthinking franchisee, such as Jim's Mowing perhaps, whereby they collect the data, trim the edges, clip it all again, and send it off to be weighed and valued. And then do it all again next time.

English academics Robert-Holmes and Bradbury cite the view that this data obsession reflects an entirely different way in which schools are being seen:

> Databases reinvent teachers and children 'into data that can be measured, compared, assessed and acted upon' and suggests that children become reconfigured as 'miniature centres of calculation'. There is a sense of young children being reduced to the school's statistical raw materials that are mined and exploited for their maximum productivity gains.

While there is a degree of overstatement in this depiction, especially given the protective, caring and resilient nature of public education in Australia we might at least heed where we currently might be headed.

And it is reasonable to observe that there are indeed dire consequences with some of the policy intentions that relate to the most recent wave of the testing agenda …

Denis Fitzgerald

In the looking glass

It comes as no revelation to note that there is a huge impetus from education authorities to move this testing juggernaut online. Concomitant with this desire is a massive edu-business motivation to gain as much command of this lucrative process which is central to their multi-billion dollar business plans. While the early stages of moving the testing online fell at the first hurdle in 2017 – the capacity of school systems to technically comply – it borders on certainty that the push will be resumed.

And while there is some potential in some elements of online assessment, including possible variegated and differentiated testing, shifting it away from a mass and simultaneous test, disconnecting the test from the current reporting fever, there is much to be done before it is possible even in limited domains.

Until such testing does not advantage those students already starting with advanced keyboard and IT skills that often correspond with existing social privilege and until all schools are equipped with the capacities and facilities to allow their students a fair and equal go in these assessments, the time is not right to move the testing behemoth online.

In its apologia for robot marking, ACARA seeks to sidestep the question of students' varying capacities and experience with computers and keyboard skills by claiming that: "There are variations in how fast and well a student can type, just as there are variations in how fast and well a student can write by hand." The speciousness of this justification is exposed when one realises that handwriting speed does not correlate with social class.

The related concern arrives with the desire of the edu-businesses to command not only the setting of these tests but also the marking of them in totality and exclusively by their corporation's computers. ACARA, which has a close and worryingly fond relationship with some of these edu-businesses, has been energetically pushing plans to move all marking online, including those sections devoted to persuasive, creative and narrative writing.

These vast global corporations, glorying in names such as Measurement Incorporated, Pearson, Pacific Metrics and MetaMetrics, claim they have the technical capacity to mark anything that is tested. Teachers are aware that simple multiple choice answers can be scanned and diced by computers but naturally this is because of the low level type of response elicited from students in such assessment items.

Yet a dollar can make a person (or a corporation) say or think or believe the most extraordinary notions. What we will soon be confronted with is the proposition that these giant companies can develop algorithms that can evaluate creative forms of expression in extended prose form.

The fact is: they cannot. Computers can identify from a pre-programmed list one or more correct

answers from a scale. They can engage in reductive counting. At the more advanced level, they can identify patterns and locate some technical errors in spelling or punctuation. They can be programmed, and here we begin to enter dangerous territory, to reward as "correct" a certain answer or answers. Their machine capacities are somewhat proficient in low order parts of the curriculum.

In all of the above, these robot markers contain finite, humanly-programmed language to "mark" and identify finite forms and modes of expression. Quite naturally, as overseas experience reveals, they can therefore be gamed as teachers prepare students to develop responses to please the robot. If you know the underlying algorithms, you can thrive. Again, the underlying feature of computer-based marking is its encouragement of skills at the lower end of the spectrum.

But a profit bottom line can be a dangerous thing and we should prepare ourselves for the march of the robots cheered on by very well paid acolytes who will swear that these machines can mark tests as well as (if not better) than professional teachers, or to use ACARA's Orwellian term "human markers".

But as Hamish Douglass, one of Australia's leading IT venture capitalists admits, we are not even close to having computer capacity to appreciate or evaluate human language and expression. As even this gung ho tech disciple admits: "We are within five to ten years away from software programs written to truly automate human knowledge and what is missing is computer understanding of natural language." Meanwhile, NAPLAN Online is being trialled in schools right now.

In the USA, whence lunacy can often emanate, robot marking has spread across school systems and into some of their colleges.

Thankfully, there has been a backlash led by academics such as Les Perelman from MIT and colleagues at Harvard and other colleges. Perelman has demonstrated how these machines can be gamed and exposed. As he explains, computer markers "do not understand meaning, and they are not sentient. They do not react to language; they merely count it".

Perelman and colleagues point out that computers are unable to recognise or judge the elements most associated with good writing such as logic, flair, clarity, accuracy, relevance of ideas, irony, innovation, playful form, appeals to audience, humour, sequencing, truthfulness, variation of tone, modes of persuasion, complex organisation, appropriate use of evidence, devices such as repetition, the deliberate use of grammatical nonconformity and sheer joyful flights of language.

The US National Council of Teachers of English concurs with the Perelman analysis. They found that computers can detect low-grade, technical errors. They can mark to a deficit system of tracking down and punishing some exam room blemishes. But little more.

They are cheaper than humans, however, and they are coming.

Denis Fitzgerald

The human costs of success

When one searches for the motivations behind much of the testing fetish of recent decades, one cannot overlook the role of politicians in barracking for greater national achievement in international tests in the quest for national pride, not unlike the parade of politicians sending our teams off to an Olympic Games. This has been facilitated by the OECD administering an Education Global Positioning System that ranks its member nations on this performance grid.

There are decades now of this scholastic cheerleading from politicians and populists, who parade a nationalist vanity and fear us falling behind our Asian competitors. Where each Asian triumph is a dagger to their heart; another Sputnik looming.

Indeed, the Federal Education Act 2013 sought to enshrine Australia's role in this educational arms race by encasing in law a national goal of being in the Top 5 PISA nations by 2025, whilst various "think-tanks" have sought to investigate how we can emulate the East Asian countries and transplant their methods to Australian schools. Meanwhile, there is increasing evidence of various countries gaming the results and seeking to extract domestic political kudos.

The recent international study by Sellar, Thompson and Rutkowski cautioned all involved to be cautious in the uses of PISA, in the interpretations of its data, in realising the chasm that can exist between causation and correlation and warned that "PISA envy is toxic".

Nonetheless, the Australian Prime Minister, in May 2017, carried on in the gung-ho tradition when he abandoned the NERA Schools Funding Agreement and dressed such abandonment in the cause of having a shiny new plan that would take us to the very top of the international testing rankings. Gold, gold, gold for Australia.

Yet when one examines the human impact on the youngsters who deliver to their nations these podium finishes, there is more than pause for thought.

Anecdotally, we have heard of excessive tutoring, high suicide rates, long days and nights of little but study and cramming for these Asian Olympians. At about the same time that Malcolm Turnbull was performing his cheerleading, the OECD released its first student wellbeing study.

The OECD examined the lives of 540,000 students across the globe. It found that those countries, especially those in East Asia, that have topped the international test medal tally have children that have paid a very high price for the success.

As Trevor Cobbald concludes from the study:

Denis Fitzgerald

> While East Asian Countries are at the top of the league table of test scores, they are at the bottom in student wellbeing …
>
> Japan, South Korea, Hong Kong, Taiwan, Macao and four mainland Chinese cities (Beijing, Shanghai, Jiangsu and Guangdong) participating in PISA have the lowest level of life satisfaction among students in 50 countries participating in the survey …
>
> The OECD data also show that students in several East Asian countries spend a large fraction of their waking hours in school lessons and studying outside of school …
>
> The OECD report raised concerns that the long hours of study by students mean less time on leisure time out of school and can come at the expense of the quality of life …
>
> One cost in many countries where students spend a lot of hours studying is that they spend little or no time in physical activity … [which] has significant implications for health and wellbeing … and these students tend to fare poorly in several psychosocial outcomes and are more likely to engage in risky behaviours … [and] about 10-20% of students at the end of school in East Asian cities are highly myopic. Early effective blindness amongst the adult population in these countries is now a major health issue with the emergence of the epidemic.

Now here is data that speaks for itself and gives us cause to reflect on what we would wish for our own students and our own children. This is an existential question only a few removes from our contemplation of testing and assessment policy.

A new road?

What we are doing as a nation in relation to assessment and related processes is causing harm – to our young, to our ethos of public education and to our futures. The evidence is in. The evidence is clear. Current practices and policies must be changed – fundamentally and expeditiously.

References and readings:

This paper has relied and drawn upon a range of academic readings, interviews, conversations and research papers on the issues it traverses.

These interactions have been with academics, teachers, principals, bureaucrats and experts with engagement in the fields of assessment, testing and method. Their contribution has been essential to the project. I deeply appreciate the assistance rendered.

What follows is a select list of readings and research that have been used as well as a range of other distinguished contributions to the considerations within the paper.

www.cpl.asn.au/journal

Denis Fitzgerald

ACARA, www.nap.edu.au/online-assessment

Paul Brock, "Show an Affirming Flame: A Message to the Profession", The Journal of Professional Learning, Semester 2 2015 at http://cpl.asn.au/journal/semester-2-2015/show-an-affirming-flame-a-message-to-the-profession

Centre for Education Statistics and Evaluation [CESE], "What works best: evidence-based practices to help improve NSW student performance" Sydney 2014 at http://www.cese.nsw.gov.au/images/stories/PDF/What-works-best_FA-2015_AA.pdf

Linda Darling-Hammond, The Flat World and Education, New York Teachers College, 2010

Andy Hargreaves and Michael Fullan, Professional Capital, London Routledge, 2010

Joanne Jarvis and Judy King, "Lifting student achievement through a culture of effective assessment in your school", Georges River College, 2017

Elizabeth Knight, "Day looms when we will be over Uber" Sydney Morning Herald May 27-28 2017

Steve Kolowich "Writing Instructor, Skeptical of Automated Grading, Pits Machine Vs Machine" in The Chronicle of Higher Education at http://chronicle.com/article/Writing-Instructor-Skeptical/146211/

G Munns, W Sawyer B Cole and the Fair Go Team, Exemplary Teachers of Students in Poverty, Routledge, London, 2013

National Council on Teaching of English, "NCTE Position Statement on Machine Scoring" www.ncte.org/positions/statements/machine_scoring and "Professionals Against Machine Scoring of Student Essays in High-Stakes Assessment", http://humanreaders.org

National Union of Teachers, The Mismeasurement of Learning, London, N.U.T., 2016

Queensland, Australia, "Rich Tasks Assessment Program", www.fairtest.org/queensland-australia-rich-tasks-assessment-program

Pasi Sahlberg and Jonathan Hasak, "Next Big Thing in Education: Small Data", reprinted in Journal of Professional Learning, Semester 1, 2017 at http://cpl.asn.au/journal/semester-1-2017/next-big-thing-in-education-small-data

Wayne Sawyer, "High Cognitive Work Across the School Years", The Journal of Professional Learning, Semester 1 2015 at http://cpl.asn.au/journal/semester-1-2015/high-cognitive-work-across-the-school-years

Wayne Sawyer,"Lifting Achievement in Years 7-12" papers and presentation to courses at the NSWTF Centre for Professional Learning [CPL]

Sam Sellar, Greg Thompson and David Rutkowski, The Global Education Race Brush Education Canada 2017

Denis Fitzgerald

Lorrie Shepard, William R Penuel and Kristen L Davidson "Design principles for new systems of assessment", Phi Delta Kappan, March 2017

Peter Wilby, "Teaching guru is optimistic about education",
https://www.theguardian.com/education/2011/jan/18/teaching-methods-government-reforms

Dylan Wiliam, "The bridge between teaching and learning", Pedagogiska Magasinet [Sweden], September 16, 2011

Jenny Williams, "Assessing Assessment K-10", in Journal of Professional Learning, Semester 1, 2016 at
http://cpl.asn.au/journal/semester-1-2016/assessing-assessment-k-10

Woonona High School, "Authentic Assessment", 2017

Denis Fitzgerald has taught in a range of public schools across NSW as a classroom teacher and in a variety of promotions positions. Denis was a foundation member of the NSW Board of Secondary Education, a member of the Curriculum Corporation of Australia and a foundation member of the NSW Board of Studies. He has also been the President of the NSW Teachers Federation and the Federal President of the Australian Education Union. Denis is currently a member of the NESA Board.