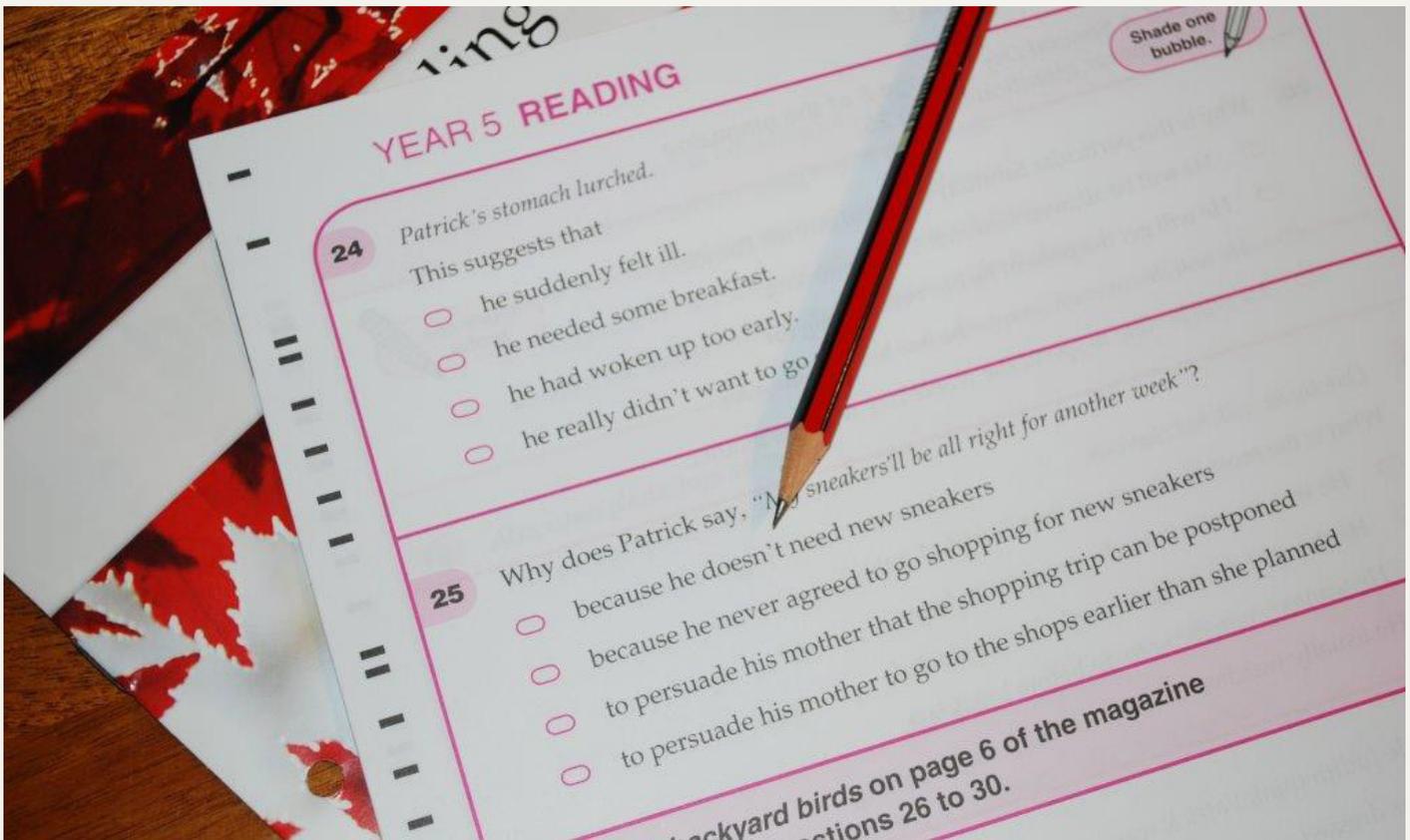


e-leading

Management strategies for school leaders

Picking up the gauntlet: Placing teacher judgement above NAPLAN



In Medieval times, a knight would challenge his enemy to a duel by throwing one of his gauntlets on the ground. The opponent would accept the challenge by picking up the gauntlet. In similar vein, Jim Watterston threw a gauntlet at the feet of delegates at last year's ACEL conference when he told the audience to stop criticising NAPLAN and to find additional measures to report to parents. NAPLAN tests, he argued, provide teachers with reliable and defensible evidence in order to monitor

progress. Watterston's point is that, rather than protesting, schools need to develop additional measures to provide greater breadth and triangulation in reporting to parents. In this context, it would be remiss if, as passionate educators, we did not pick up Watterston's gauntlet and accept the challenge, however onerous, of finding an alternative to NAPLAN.

Governments invest in large-scale assessment programs for two reasons. Firstly, because of an expectation that educational systems should be

accountable for student achievement and, secondly, that parents and the community should have access to comparable information about student performance. Large-scale testing programs rarely, however, sit comfortably with educators even where significant efforts are made to help teachers use the data to reflect on their teaching. Considerable work is currently underway internationally to find solutions that enable schools to be accountable for outcomes in ways that are both valid and reliable. Policymakers



e-leading July 2014 (22) – Researched and prepared for ACEL by Dr Sandy Heldsinger, Director of Pairwise Pty Ltd, Educational Consultant and Researcher and Dr Stephen Humphry, Associate Professor of Educational Assessment and Measurement, University of Western Australia.

Also published by ACEL e-shortcuts – Wisdom for successful school leadership and management & e-teaching – Management strategies for the classroom. *Subscribe online at www.acel.org.au*

Table 1: Sample of Pairwise studies showing the separation indices

Assessment Area	Separation Indices
Early writing development (K – year 2)	0.99
Narrative, persuasive and informational report writing (Years 1 – 7)	0.95–0.97
Science investigation (years 4 and 5)	0.96
Oral development (K – year 4)	0.95
Violin	0.91
English (Creative writing)	0.95
English (Narrative and Essay)	0.97
Visual Arts	0.95
Philosophy	0.84
Accounting and finance	0.95

in the USA are considering formative assessment as a primary approach to educational reform and there has been considerable investment in standards-based assessment systems. In England, several models are presently being explored and include: a sampling model to moderate teacher assessments; a combination of school inspections and teacher assessment providing information to parents and children and adaptive, on-demand testing.

For many in education, the ideal solution would be to harness teacher judgements to monitor and report student outcomes. Advocates of such an approach however are not only cognisant that obtaining reliable judgements is an enormous challenge, but also that such a system could add significantly to the pressure on teachers. “There would most likely arise significant tensions relating to the forms of data that would be required by government, particularly between national data and accountability demands and the kind of teacher assessment that teachers would most likely desire as an optimum approach for supporting teaching and learning”¹

To find an alternative to NAPLAN, we must first appreciate the strengths of the program. NAPLAN enables teachers and schools to evaluate their impact on student learning because it provides comparative data across classrooms and schools. There is another reason, one which is not explicitly acknowledged, but is perhaps more important. The NAPLAN assessments are equated from one round of testing to the next, and from one year level of testing to another. Schools can therefore use the data to evaluate improvement from one year to the next and as a result they can also evaluate student growth in learning. Whilst teacher anecdotal records and teacher devised tasks often provide great insights into students’ skills and understandings, these assessments can only be used to roughly gauge the progress a student has made or the impact a program has had on student learning. It is necessary to use a test where a measurement scale has been established to determine – with some certainty – how much progress students have made or the impact of a program on learning.

The challenge, however, is whether it could ever be possible for teacher judgement to be at the heart of school accountability? In other words, could teacher judgements give schools the same (or even better) quality of data than they currently get from NAPLAN?

Excitingly, yes we think it’s possible. We do not offer this solution lightly. The breakthrough is a result of a decade of pioneering research at the University of Western Australia. The

assessment method developed from this research is intuitive and involves teachers comparing pairs of performances; the teachers judge which performance, in each pair, demonstrates more ability in terms of the construct being assessed.² We have studied the methodology in over 100 studies in diverse aspects of the curriculum and across the full educational spectrum including early childhood, primary, secondary and tertiary education. In virtually all studies one constant has shone through: we have found that teachers are highly consistent in judging relative differences of student ability. Table 1 lists a sample of the studies conducted over the past ten years at UWA. The separation index provided in the table is an indicator of consistency or reliability where 1.0 indicates perfect agreement across the judges and anything over 0.9 shows very high agreement. The separation index can be compared to the reliability coefficient reported in standardised testing programs and as a point of reference, the reliability for NAPLAN reading and numeracy is traditionally around 0.85–0.9. Across nearly all our studies, including these, we have found that where the pairwise comparison methodology is used to capture teacher judgements, those judgements are as reliable, if not more reliable, than NAPLAN data.

While the method of pairwise comparisons provides reliable teacher judgements, it is not necessarily a particularly efficient way for teachers to assess. Hence, we have investigated a two-stage process where the pairwise method is used to calibrate performances and develop a performance scale. Teachers then assess their students’ work simply by judging which of the calibrated performance a student’s work is most like. We have found that this two-stage process is efficient and provides highly reliable teacher judgements of student achievement. The methodology is summarised in Figure 1.

To assess and score a student’s work, a teacher compares a student’s performance to the calibrated exemplars and decides which exemplar the performance is most similar to, in terms of level of performance, or which two exemplars it falls between. The performance descriptions help teachers gauge the range in the scale where they can begin to compare their student’s work to the calibrated exemplars. Once satisfied with his or her judgement, the teacher simply locks it in. The student’s score is captured and the teacher can move on to assess the next performance.

There are two features of this assessment process that makes it a powerful assessment tool. The process leads to reliable and consistent teacher judgements. It also results in all students

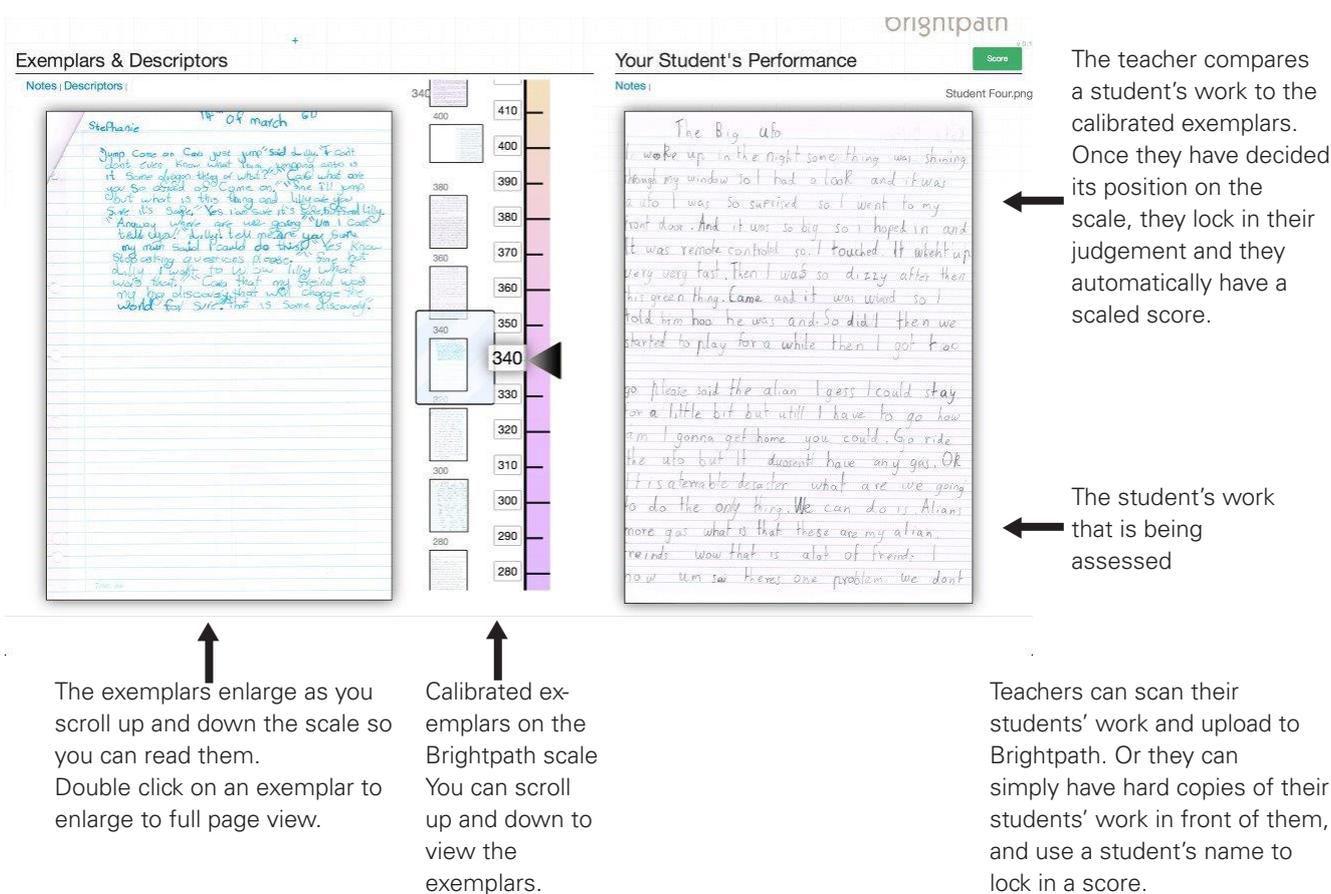


Figure 1: The Teacher's Ruler

being placed on the same scale (for a given assessment) regardless of their year level.

Essentially, this means that data collected from teacher judgements can be used to evaluate school programs in much the same way that NAPLAN data can be used. School leaders and teachers can validly review the spread of student abilities within a class, across classes and across cohorts. After two or more rounds of assessment, schools can use the data to evaluate student growth in learning.

The Western Australian Primary Principals Association and the Australian Primary Principals Association are now leading an initiative to make the methodology readily accessible to all primary schools, in the first instance. Readers can find out more about this exciting development by going to www.brightpath.com.au. We believe that teachers can pick up Watterston's gauntlet, by combining their professional knowledge of their students with the pairwise comparison assessment model. Very soon an accountability system that relies solely on externally imposed standardised testing such as NAPLAN may well be seen as something from the dark ages.

Endnotes

- Green, S. & Oates, T. 2009, 'Considering alternatives to national assessment arrangements in England: possibilities and opportunities', *Education Researcher*, vol 51, no 2: 229-45.
- Heldsinger, S. & Humphry, S. 2010, 'Using the method of pairwise comparison to obtain reliable teacher assessments', *The Australian Educational Researcher* vol. 37, no 2: 1-19.

Quotes on being a professional

Once you realize that you're in something that you've always wanted and you don't want to lose it, you behave differently. And that means the integrity, the professionalism, and knowing what's right from wrong and still making choices that you probably wouldn't have made.

Paul Anka

I believe in professionalism, but playing is not like a job. You have to be grateful to have the opportunity to play.

Wynton Marsalis

Professionalism is not sportsmanship. If you don't succeed, you won't be in your profession for long. In our society, it's not about good or bad. It's about who's on top.

Chili Davis

I think professionalism is important, and professionalism means you get paid.

Erica Jong