

Student Knowledge: Curriculum, Assessment and Reporting

Maria Nicholas¹

Abstract

At a time when national and international high-stakes testing has assumed such prominence, one might begin to wonder about the status of teacher judgement when assessing and reporting on children's knowledge and skills against the descriptors specified in curriculum standards. Were standardised test results congruent with the judgements that teachers make when reporting on students' achievement, concern about how one type of judgement might compare with another would perhaps be unwarranted. This article draws on research that has investigated whether standardised assessments in the state of Victoria, Australia are actually comparable with teacher's judgements about their students' work to illustrate that discrepancies do exist. These results have been interpreted within an analytical framework that derives from Aristotle's (350BC/2000) distinction between three types of knowledge, namely epistemic, technical and phronetic knowledge.

Key words: Teacher judgement, Curriculum, Assessment, Reporting

Introduction

At a time when national and international high-stakes testing has assumed such prominence, one might begin to wonder about the status of teacher judgement, both the judgements that teachers reach in their day-to-day interactions with children and their estimations of children's abilities against the detailed outcomes specified in curriculum standards. Were standardised test results congruent with the judgements that teachers make when reporting on students' achievement, concern about how one type of judgement might compare with another would perhaps be unwarranted. This article, however, which draws on research that has investigated whether standardised assessments in the state of Victoria, Australia are actually comparable with teacher's judgements about their students' work, illustrates that discrepancies do exist.

As part of this study, teacher judgement of Year 2 and Year 3 students' Reading outcomes were collected from 42 government schools from across the state. These results were collected along with students' Year 3 NAPLAN (National Assessment Program – Literacy and Numeracy): Reading scores. These results have been interpreted within an analytical framework that derives from Aristotle's (350BC/2000) distinction between three types of knowledge, namely epistemic, technical and phronetic knowledge (see also Biesta, 2013; Kemmis, 2015 and Carr, 2005).

Curriculum

Curriculum has traditionally been defined as "a knowledge structure – a statement about what is 'to count' as learning in a given domain of enquiry and about how learning should progress during a nominated period of apprenticeship" (Macken-Horarik, 2011, p. 197). The creation of such statements can be a contentious endeavour, as teachers, teacher educators and academics may not all agree on what is to count as "valued knowledge" in any given field (Macken-Horarik, 2011; Victorian Curriculum and Assessment Authority, 2014). In this sense, curriculum can be said to be conjectural. It is a depiction of its creators' expectations and theories on the 'when' and 'what' of knowledge and skill acquisition. It operates within an arena of

¹ Deakin University, corresponding author maria.n@deakin.edu.au

generalisations, “as if” (Sfard, 1998) the mind were independent of the world (Packer & Goicoechea, 2000). These expectations and theories are ideological in nature and contextualised, as highlighted by the content of Victoria’s curriculum in the 1880s which included the teaching and learning of needlework for female students (Sweetman, Long, & Smyth, 1922).

Working within this “knowledge structure”, Yates and Collins (2010, p. 89) contend that over the past few decades a significant shift has occurred in Australian curriculum policies, moving from placing emphasis on “knowing things to being able to do things”. It was in the 1980s, with the first attempts at introducing a National Curriculum that contained “outcomes” as well as “profiles”, that Australia first attempted to marry “a child-focused (i.e. cognitive developmental) view of curriculum to the instrumental outcomes-based desire of those steering the national agenda” (Yates & Collins, 2010, p. 92). Rather than being presented with a curriculum that *prescribed* the ‘when’ and ‘what’ of knowledge acquisition, teachers had been provided with a developmental “continuum of learning” (ACARA, 2013b) that placed greater focus on their students’ cognitive skills and outputs, and on their own professional judgement and delivery. This change necessitated them knowing their students individually, knowing where each of them lay along that continuum and how they could deliver content through meaningful processes.

Assessment and Reporting

Assessment is a powerful, multipurpose tool. It enables teachers to track student progress with an eye towards informing their planning, teaching and reporting. In addition to this, it also has the potential to place a spotlight on the teacher when seeking to ensure reliability, accountability, transparency and consistency across contexts, and to positively contribute to public confidence in the education system (Klenowski, 2012; Klenowski & Wyatt-Smith, 2012). In the state of Victoria, the curriculum with its levelled achievement standards, is to be used by teachers to inform the reports they provide to parents and students each year, detailing their progress against common achievement standards “including an indication of student progress against the age-related expected level of achievement” (Victorian Curriculum and Assessment Authority, 2014, p. 6). It is this particular use of the term ‘reporting’ which will be used within the context of this article.

Though schools are to make use of the curriculum to report against student standards of achievement, it is acknowledged that “schools have considerable flexibility in the design of their teaching and learning program” (Victorian Curriculum and Assessment Authority, 2014, p. 5) and in the way in which student outcomes will be reported to parents and to the students themselves (Department of Education and Training, 2014; Victorian Curriculum and Assessment Authority, 2014). Given this allowance for flexibility, for both the student and the professional who is given the task of interpreting the standards, interpretation of the standards and how they are used to inform assessment and reporting can prove problematic. In a study that saw teachers engaged in moderation sessions in the state of Queensland, assessing students against curriculum achievement standards, participants reported the sessions to be difficult and consistency of teacher judgement to be elusive (Connolly, Klenowski, & Wyatt-Smith, 2012).

Teachers stated that assessing students from one context against the same achievement standards used with students from a different context can be difficult, as those children will not have based their work on the same prior knowledge, skills, experiences and perceptions of the other (Connolly et al., 2012). Teachers also reported that the accuracy of their assessment was impacted by the “perceived ambiguity” and “lack of

detail” of standards descriptors (Connolly et al., 2012, p. 13). This would suggest that when working with the generalised wording of a curriculum, assessment of student work samples and the allocation of levels or grades for the purpose of reporting can become more a process of deciding on the “best fit”, than on finding the “perfect match” (Klenowski, 2012, p. 93). That said, reporting on student achievement by its very nature necessitates a trust in the professional, reflective judgement of the teacher; a practitioner who is uniquely privileged and in a position to appreciate and account for the many factors that may have contributed to each child’s practice over time. It could even be argued that this is one of the main qualities that speaks to a teacher’s efficacy, as many teachers judge their own professionalism on their ability to recognise and appreciate the diversity of all of their students, considering this to be a key indication that they themselves are engaging in “good practice or true education” (Kostogriz & Doecke, 2011, p. 407). Klenowski (2012) describes this type of longer-term, multifaceted approach to reporting as an “authentic assessment” as these assessments are seen to reduce:

dependence on performance in a single terminal examination as the only determinant of student achievement... [to give] individuals the opportunity to demonstrate attainment over time and in a variety of contexts. This type of assessment is claimed to be more accurate, and reflective of an individual’s learning and development... [and] encompass a wider range of abilities. (p.95)

It would have been of interest to determine if consistency of teacher judgement would have been easier to achieve, had the teacher participants of Connolly and colleagues’ (2012) study been asked to assess the work of students they had all worked with and come to know over time.

It has been suggested that student assessments first attracted attention at a time when schooling, assessments, and higher education became seen as avenues through which to enact a change in social class, following on the trails of the industrial revolution (Earl, 2005). Authors such as Earl (2005) and Klenowski and Wyatt-Smith (2012) submit that it was this appreciation that resulted in a perceived need to monitor the “gatekeepers” (Earl, 2005, p. 5); those who were seen to be in a position to be able to decide who would progress through, who would be deemed as having ‘potential’ and who would not. It has been suggested that it was this perceived need for objectivity that sparked the introduction of the standardised, scientifically-measured assessment at the local level. It could also be argued that it was in recognition of the desire to accurately award students comparable levels of achievement and grades that saw the uptake of standardised assessments as a useful tool to complement and help facilitate more accurate and informed teaching and reporting.

The Epistemic, the Technical and the Practical

...the mere handling of any other instrument will never make any one a true craftsman or athlete, nor will such instrument be even useful to one who has neither learnt its capabilities nor exercised himself sufficiently in its practical applications.

(Plato, 380BC/2013, p. 126)

The analytical framework which has informed this study can be traced back to the work of Aristotle. Influenced by the teachings of Plato, Aristotle surmised that knowledge can be classified into three distinct categories: the epistemic (theoretical), the technical (or productive) and that of experience (phronetic knowledge). He surmised that the epistemic is “knowledge of universals”; that technical knowledge is “knowledge of causality”, knowledge of the underscoring “why”; and that from experiences or practice

comes “knowledge of the individual” (Aristotle, 350BC/2000). These distinctions have continued to hold sway with epistemic reasoning and technical knowledge of what *can* be done, considered by some to differ from the practical, which necessitates the subjective use of judgement in deciding what *should* be done through reflection (Biesta, 2013; Carr, 2005; Kemmis, 2005). As suggested by Plato’s introductory quote however, these distinctions are accompanied with the suggestion that one who can function within the integrated web of all three knowledges is superior in knowledge to one who is expert in only one.

In reference to student knowledge, assessment and reporting, curriculum achievement standards can be said to identify what their author/s theorise that typically developing students should know at each stage of development. The standardised assessment, a decontextualised list of student competencies insofar as they does not reflect the individualised nature of students’ everyday lives, can be said to identify the technical knowledge that students have been capable of demonstrating at the time of assessment. Teacher judgement of student achievement, reflecting on what students have demonstrated over time as part of their everyday school based experiences, can be said to identify what students can do in practice.

The Privileging of the Standardised Assessment

In the current climate, it would appear that heightened import is being placed on comparative assessments such as the Program for International Student Assessment (PISA). It has been proposed that it is this move towards a competitive focus on comparability that has seen governments the world over moving towards the implementation of standards-based curriculum and national assessments (Connolly et al., 2012). In recent times, the *My School* website was created and released in Australia. This site was designed with the intention of representing “a new level of ‘transparency’ that enables parents to locate statistical and contextual information about schools, see how they perform, compare them and supposedly make informed choices” (Kostogriz & Doecke, 2011, p. 400). Yet the sole assessment outcomes listed on this site to date have been those of the National Assessment Program – Literacy and Numeracy (NAPLAN), privileging these standardised assessment results over the grades given by teachers when reporting against the achievement standards listed in curriculum documents. This has received criticism, with a suggestion that such a focus increasingly blurs the “distinction between managerial accountability and professional responsibility” (Kostogriz & Doecke, 2011, p. 400).

When value is only placed on the performance of a stand-alone, standardised assessment, such as the NAPLAN, it can be argued that there is an accompanying suggestion that a child’s technical knowledge (what they ‘can’ do) holds more value than their practical knowledge and skills (what they ‘do’ in context). Such a focus “imposes on teachers a new form of accountability that is perceived by teachers as detached from both their situated practice and the needs of children and parents” (Kostogriz & Doecke, 2011, p. 403). Some have even pointed to “unintentional consequences” such as

the narrowing of the curriculum as teachers teach only that which is to be tested; curriculum areas that are not tested are neglected; higher order thinking skills that are difficult to assess in such paper and pencil formats are also neglected; time is spent on coaching and practice tests; schools participate in perverse practices designed to improve achievement data.
(Klenowski & Wyatt-Smith, 2012, p. 70)

These “unintentional consequences” aside, were teacher judgement outcomes to perfectly align to the outcomes of standardised assessments, the privileging of standardised assessments would not prove overly problematic. Standardised assessments could be seen as an objective data source, used to support, reflect and

reinforce the subjective, situated judgement of the teacher. Yet, the teachers referred to earlier (Connolly et al., 2012; Doecke, Green, Kostogris, Reid, & Sawyer, 2007), and others, (Freebody & Wyatt-Smith, 2009; Mercurio, 2005) have spoken to a ‘mismatch’ between teacher judgement and standardised curriculum descriptors or assessments, pointing to difference between theory and situated practice, between purpose and interpretation, and between intention and outcome, as potential sources. An alignment may also be difficult to action given the perceived differences in reasoning that can be attributed to the identification of the epistemic (curriculum), the technical (standardised assessment) and the practical knowledges (the situated practice of ‘doing’) under consideration. The intention of this paper is to provide insight into the question of comparability using the following line of inquiry:

Are teacher-based assessments of children’s overall skills and knowledges comparable to standardised assessments, which report on children’s technical skills alone?

The Current Study

The current Australian Curriculum for *English, Mathematics, Science, Humanities and Social Sciences* was preceded in the state of Victoria by the Victorian Essential Learning Standards (VELS), which was introduced in 2005. At the time of data collection (2011-2012), Grades 2 and 3 primary school teachers were reporting against the VELS when awarding students end of year grades. Much as is the case with the Australian Curriculum, the VELS listed “standards” for each level of development (levels 1 to 6), designed to “outline the essential knowledge, skills and behaviours students [were] expected to demonstrate within each domain [learning area]” (Victorian Curriculum and Assessment Authority, 2009). This was accompanied by the state government’s creation of online support documents: the *English Developmental Continuum P-10* (DEECD, 2013a), the *Mathematics Developmental Continuum P-10* (DEECD, 2013b) and the *Science Continuum P-10* (DEECD, 2013c), highlighting the intention that teachers make use of the curriculum as an overview of the continuum of learning, rather than as a year by year syllabus.

This continuum could also be used as a benchmark to determine whether each child was operating at an expected level, above expected or below expected when compared to their peers. By the end of the first year (the preparatory year), a typically developing child could expect to have demonstrated that they were working at a level 1 standard of development. Following from this, typically developing children had 2 years to progress to the next curriculum level. This meant that the typically developing Grade 1 child could be placed at level 1.5 by the end of the year; that a typically developing Grade 2 child could be placed at level 2 by the end of Grade 2; the typically developing Grade 3 child could be placed at level 2.5; that the typically developing Grade 4 child could be placed at level 3 by the end of Grade 4, and so on. The standards statements assisted teacher to determine whether each child had demonstrated what was expected of students working at the various levels of ability. For example, a Level 2 standards statement read:

They locate directly stated information, retell ideas in sequence using vocabulary and phrases from the text, and interpret labelled diagrams.

(Victorian Curriculum and Assessment Authority, 2012, np, online)

As detailed above, teachers were free to use a variety of methods to determine whether children had demonstrated the ability to meet the various achievement standards, and ultimately the grade they awarded each child.

To determine whether teacher based assessments of children’s overall skills and knowledges are comparable to standardised assessments which report on children’s technical skills alone, Grade 2 teacher judgements

(VELS: Reading), Grade 3 teacher judgements (VELS: Reading) and Grade 3 NAPLAN: Reading results were used to inform this study, as reported to the Department of Education and Early Childhood Development.

Participants

The participants represented in this study included 1,591 Year 2 students enrolled in 42 Victorian schools in December 2011, 1,673 Year 3 students from those same schools who were eligible to participate in the NAPLAN Reading assessment in May 2012, and the 1,570 Year 3 students enrolled in those same schools in December 2012. The schools that volunteered to take part in this study were located in all four regions from the state of Victoria: 15 schools were from the North-Western region, 8 from the North-Eastern region, 10 from the South-Western region and 9 schools were from the South-Eastern region.

The Australian Curriculum, Assessment and Reporting Authority (ACARA) has calculated the *Index of Community Socio-educational Advantage* (ICSEA) values for Australian schools for each calendar year from 2008 to 2014, and has published these values on its *My School* website. In 2012, the formula (“in broad terms”) was calculated as follows:

ICSEA = SEA [socio-economic advantage] (direct/indirect) + Remoteness + Percent Indigenous student enrolment
(ACARA, 2013a)

“Direct data” was categorised as parent background information, provided upon enrolment and via NAPLAN reporting (ACARA, 2013a, p. 1).

41 of the 42 schools used in this study were provided with an ICSEA value in 2012. Grounds upon which schools might not have been provided with ICSEA values ranged from the school being categorised as a specialist school or a school with insufficient “aggregate level data” (ACARA, 2013a). Of the 41 participant schools with available data, ICSEA values ranged from 928 to 1175 with an average of 992; 21 schools scoring above 1000 and 20 scoring below. The median value for all Australian schools as provided by ACARA is 1000, with values ranging:

...from around 500 (representing extremely educationally disadvantaged backgrounds) to about 1300 (representing schools with students with very educationally advantaged backgrounds).
(ACARA, 2013a)

Though the reliability of the formula used to calculate ICSEA values was under review in 2013, in the absence of other more reliable data, this data would suggest that on average, most of the students from the participant schools were representative of students who came from backgrounds that were neither highly educationally advantaged nor extremely disadvantaged.

Given the non-identifiable nature of data collection, it cannot be known how many of the 1,591 Year 2 students who were represented in the 2011 data were the same students represented in the Year 3 2012 data. It also cannot be known how many of the 1,570 students represented in the Year 3 ‘VELS teacher judgement’ category, were represented in the Year 3 ‘NAPLAN’ data set which represents the 1,673 students who were eligible to take part in the NAPLAN reading assessment at those 42 schools in 2012.

Data provided by the Australian Bureau of Statistics (2012) indicates that between 2006 and 2011, 41.7% of people who were aged 5 years and over (7.6 million) changed their place of residence across the country. Of the 4.6 million longer-term residents, 37% lived in a family with children under 15 years. Of the 3 million 'new residents' (1.8 million who moved from a different area in Australia, and 1.2 million who were overseas in 2006), just over 39% were in a family with children under 15 years of age. This equates to 2,872,000 residents, or 38% of the population over a 5-year period, which on a yearly average equates to 7.6% of individuals who lived in a family with children aged under 15 years. These statistics would support the assumption that approximately 92.4% or more of the participants involved in this study were represented across all three samples.

Method

All Victorian government primary schools (of which there were 1260 in 2013) were invited to take part in this study. 42 schools elected to take part and gave consent to have their Department of Education and Early Childhood Development (DEECD) owned non-identifiable data released to the researchers. Though all schools in Victoria report against and use the same curriculum, schools had and continue to have (under the renamed Department of Education and Training) some flexibility in what is included in their teaching and learning programs and how they report student outcomes to students and to parents (Department of Education and Training, 2014; Victorian Curriculum and Assessment Authority, 2014). To allow for comparability, only DEECD owned data sets (VELS levels and NAPLAN data) were sought for this study.

These data were the grades (VELS: Reading) and results (NAPLAN: Reading) on record on the DEECD database. The researchers received the following data from the DEECD:

- December, 2011 Year 2 teacher judgement VELS (Reading) data for each student from each of the participant schools (the number of students who received an A grading, the number who received a B grading, the number who received a C grading, a D grading and an E grading).
- December, 2012 Year 3 teacher judgement VELS (Reading) data for each student from each of the participant schools (the number of students who received an A grading, the number who received a B grading, the number who received a C grading, a D grading and an E grading).
- May, 2012 Year 3 NAPLAN Reading data for each student who participated from each of the 42 schools (the number of students who fell within band 1, the number who fell within band 2, band 3, band 4, band 5 and band 6).

Student names and gender were not provided.

In acknowledging the issues that arise from the development of a curriculum, which some argue is immediately outdated upon creation and others submit is missing critical content essential to life-long learning due to the vast number of contenders which could feasibly have been included at the various levels of development, the VCAA published the following statement about the development of curriculum:

it is essential for educators to define a *minimum* [emphasis added] and limited set of declarative and procedural knowledge and skills that all students should acquire, irrespective of their personal inclinations. (Victorian Curriculum and Assessment Authority, 2014, p. 8)

This would indicate that level 2 was the minimum standard expected of a student who had completed Grade 2, that level 3 was the minimum standard expected of a child who had completed Grade 4, and that level 4 was the minimum standard expected of a child who had completed Grade 6 under the VELS.

NAPLAN guidelines state that:

...a national minimum standard is defined and located on the assessment scale for each year level. Band 2 is the minimum standard for Year 3, band 4 is the minimum standard for Year 5, band 5 is the minimum standard for Year 7 and band 6 is the minimum standard for Year 9.
(Australian Curriculum, 2013, np, online)

Using these definitions, we are better placed to compare the 5-point scale of the VELS against the 6-point scale of NAPLAN.

Results

Table 1, Table 2 and Figure 1 list the number of students who were awarded particular grades by their teachers for the Victorian Essential Learning Standards (VELS) domain of Reading in Year 2 of 2011 and Year 3 of 2012. They also list the number of students who were awarded particular “bands” following their Year 3 NAPLAN Reading assessment in 2012.

The VELS “C” grading signifies that the student was deemed by their teacher to be working “at expected” level. “D” signifies that the child was deemed to be working a year below expected. “E” signifies that the child was deemed to be working more than a year below expected level. “B” signifies that the child was deemed to be working six months ahead of the expected level. “A” was representative of the students whom teachers believed were working one year or more above expected level. The NAPLAN “band 2” grading signifies that based on the one stand-alone standardised assessment, the student was deemed to be working at the national minimum standard for Year 3 students. “Band 4” typically signifies that the student was working at the national minimum standard for Year 5 students. “Band 1” signifies that the student was working below the national minimum standard for Year 3 students. N/A is representative of the number of enrolled students who were not assessed. It cannot be known, from the data provided, why those students were not assessed.

	E	D	C	B	A	N/A	Total
Year 2 Dec 2011 VELS:R	15	71	789	450	233	33	1591
	0.9%	4.5%	49.6%	28.3%	14.6%	2.1%	
Year 3 Dec 2012 VELS:R	16	95	776	380	281	22	1570
	1%	6.1%	49.4%	24.2%	17.9%	1.4%	

Table 1: Number of students awarded Grades E-A for VELS Reading

	Band 1	Band 2	Band 3	Band 4	Band 5	Band 6	N/A	Total
Year 3 May 2012 NAPLAN:R	55	153	289	361	323	353	139	1673
	3.3%	9.1%	17.3%	21.6%	19.3%	21.1%	8.3%	

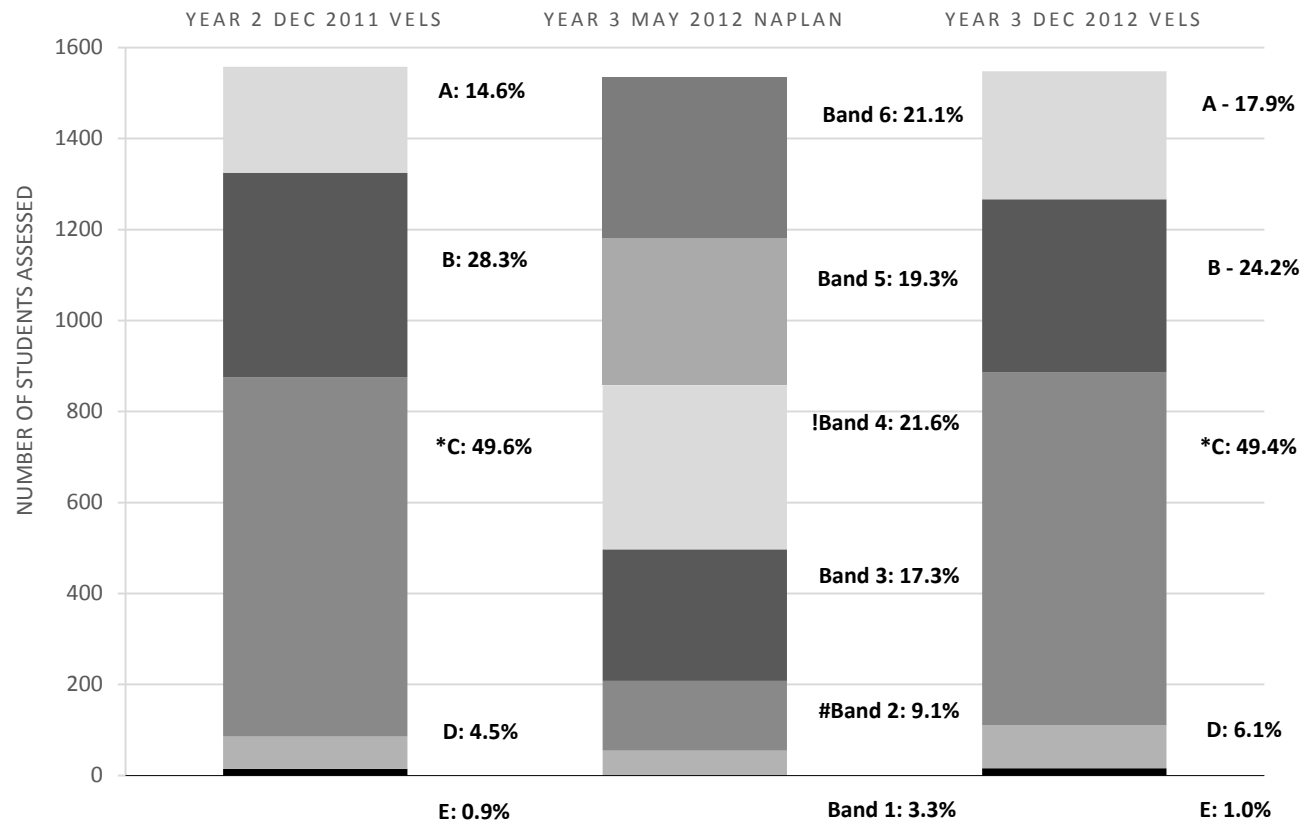
Table 2: Number of students awarded bands 1-6 for NAPLAN Reading

Figure 1 is a graphic representation of the data presented in Tables 1 and 2. It should be noted when analysing these data, a direct comparison of VELS and NAPLAN was somewhat difficult to achieve given the 5 point scale of the VELS and the 6 point scale of the NAPLAN. Comparability was achieved by using the minimum standards to guide analysis.

These data suggest that the participants’ teachers did not judge their students to be as evenly spread in ability as the NAPLAN Reading data seems to propose. The results suggest that teachers believed that a greater concentration of students (half of the students in both Year 2 and Year 3) were working at the one level (at expected level) in both of those years. The NAPLAN data however shows that only 9.1% were deemed to be working at the national minimum standard for Year 3 students when they were in Year 3. There also seemed to be a greater spread of ability within the NAPLAN results when working above the national minimum standard, with 17.3% of students deemed to be working at band 3, 21.6% at band 4, 19.3% at band 5 and 21.1% at band 6.

These data also seem to suggest that when it comes to teacher judgement for Reading, there was only marginal difference between one year and the next with a similar percentage of students working at each level of ability as they moved from the end of Year 2 to the end of Year 3. This is especially noticeable when one considers those who were deemed as working more than one year below expected level (there was only 0.1% difference between the number of students given an “E” grading in Year 2 and an “E” grading in Year 3) and those who were deemed as working “at expected level” (there was only 0.2% difference between the number of students given a “C” grading in Year 2 and a “C” grading in Year 3). The greatest difference occurred at the “B” level (6 months above expected), with a difference of 4.1%.

TEACHER JUDGEMENT AND NAPLAN READING RESULTS



*C – The number of students whom teachers deemed were working “at expected” level.

#Band 2 – These students were performing “at the national minimum standard” for Year 3 students.

!Band 4 – These students were performing “at the national minimum standard” for Year 5 students.

Figure 1: Teacher Judgement VELS: Reading and NAPLAN: Reading

Overall however, when pooled together, in 2011, 683 students (42.9%) were deemed by their teachers to be working above expected level (levels “A” and “B” combined), with 661 (42.1%) deemed as working above expected level in 2012. With a variance of only 0.8% between one year and the next, the difference is minimal.

An interesting finding is noted when one considers that the number of students represented in NAPLAN bands 2, 3 and 4 as a whole (803 students) was similar to the number of students who achieved a level “C” grade for VELS in 2011 (789 students) and in 2012 (776 students). While the VELS data represents 49.6% of the Year 2 students and 49.4% of the Year 3 students who were working “at expected” level, the 803 students placed at NAPLAN bands 2, 3 and 4 combined, represent 48% of the Year 3 students who were eligible to participate in the Year 3 NAPLAN Reading assessment; ranging in ability from the “national minimum standard for Year 3 students” right through to students who are typically at the “national minimum standard for Year 5”.

Discussion

There have been researchers into curriculum (e.g. C. Harris & Marsh, 2005), who have highlighted the problems associated with translating theory (curriculum policy) into practice (the daily work, the ‘doing’, of schools, teachers and learners). In the discussion that preceded the present study’s results, the author put forward a case for considering that an alignment may be difficult to action given the perceived differences in reasoning that have been attributed to the identification of epistemic knowledge (such as curriculum), the technical (such as standardised assessment) and the practical (the situated practice of ‘doing’). To explore whether standardised assessments may be comparable with teacher reports on student outcomes, a study was devised with the intention of comparing VELS Reading teacher judgement outcomes against the standardised assessment outcomes of Year 3 students, in the state of Victoria, Australia.

The study found that in 2011 and 2012, VELS Reading teacher judgement outcomes did not align with the standardised assessment outcomes of the Year 3 NAPLAN: Reading, of 2012. In general, teachers believed that close to half of their students (49.6% in 2011 and 49.4% in 2012) were working “at expected” level, whereas the NAPLAN: Reading results seemed to suggest that the Year 3 students from the participant schools had a more varied spread of ability. The NAPLAN results suggested that only 9.1% of students were working “at the national minimum standard” (band 2) for Year 3 students, with 17.3% of students deemed to be working above the minimum standard at band 3, 21.6% at band 4, 19.3% at band 5 and 21.1% at band 6.

This study also suggests that when students are assessed by teachers who know them, know their backgrounds and have seen their everyday work over a long period of time, teacher reports can be relatively reliable given the consistency documented between one year and the next. The present study made use of data that were obtained after the fact, without any interference (intentional or otherwise) from the researchers. In past studies, when teachers were asked to agree on a grade level for the same piece of work, consistency was difficult to achieve (Connolly et al., 2012; Cooksey, Freebody, & Wyatt-Smith, 2007; Wyatt-Smith, Klenowski, & Gunn, 2010), with Cooksey et al. (2007) finding that teachers could agree to a greater degree on the ranking of work than on precise grade levels. However, the present study would seem

to submit, that when judging the outcomes of their own students, at least in the state of Victoria, there appears to be more consistency between one year and the next than research into moderation sessions in Queensland would suggest. This might point towards the fact that when teachers can make use of their knowledge of context, consistency of judgement is easier to achieve.

Freebody and Wyatt-Smith (2009) suggest that when it comes to assessment it needs to be acknowledged that practitioners operate in two differing contexts; that of “system” and that of “site”. The first is concerned with underlying general sub-competencies that can be used for system-wide comparability purposes, whereas the latter is concerned with individualised, practical application of literacy practices in context and is therefore more difficult to neatly define or compare. A focus on the “system” through standardised assessments such as the Australian NAPLAN defines literacy for example, “as a set of technical skills... as instrumental activities with little attention paid to the contexts in which the literacy practices take place” (Snyder, 2008, p. 205). This is considered by some to be problematic as a focus on “the system” runs the risk of encouraging a narrow definition of “literacy”, while a focus on the “site” leaves government and policy makers at a loss as to how and where to focus their support efforts (Freebody & Wyatt-Smith, 2009).

When considered together, and with additional appreciation for the fact that teacher judgement involves a process of subjective interpretation (both of the standards and of student practice) one comes to appreciate why assessment has been described as “intrinsically inexact” (Klenowski, 2012, p. 91). Freebody and Wyatt-Smith (2009, p. 32) suggest that this presents educators with a significant dilemma:

They must report at levels of abstraction and generality ...[o]r, they must attend to and respect particularity at the level of theory, pedagogy and data collection – at the potential expense of system-wide comparability...

For the teacher, this becomes an issue of fairness and equity, of confidence in the fact that every student has been given the opportunity “to reach and demonstrate their current capability” (Klenowski, 2012, p. 96). Given that “[s]tudents may demonstrate their knowledge, skills and understanding in a variety of ways” (Klenowski, 2012, p. 96) the use of a stand-alone, seemingly “objective” assessment, detached from authentic practice, can prove difficult. Some would argue that this may also be due to the limited scope of the “once off” assessments to assess a child’s content knowledge and skill, limiting its usefulness to inform planning for teaching and learning (Harris et al., 2013).

But why must teachers choose one or the other?

In addressing the question that was raised earlier, the results of this study have suggested that student outcomes from a standardised assessment such as the NAPLAN aren’t always directly aligned with teacher judgement outcomes when reporting against a standardised curriculum. What this suggests is not necessarily that one is of more value than the other, but that when considered alongside Aristotle’s concepts of epistemic, technical and practical knowledge, that rather they should be awarded equal standing for differing reasons.

When comparing a standardised assessment of technical knowledge such as the NAPLAN against the assessment of practical skill as reported by teachers, comparable marks would suggest that a child had demonstrated an ability to transfer a knowledge of what *can* be done into the practical. They have demonstrated an ability to use this knowledge to inform their actions as they make judgements on what *should* be done. Lower marks in the former when compared to the latter may indicate that a child *can* through “natural tendency”, through “habit” (Aristotle, 350BC/2000), minus a deeper understanding of the theoretical

and/or the “why”. Such an outcome might indicate that these particular children may find it difficult to apply generalisable rules and knowledge to new and unfamiliar experiences and contexts. Subjective reasoning may be hampered due to a not yet fully developed understanding of universals and of the “why”. Higher marks in standardised assessments when compared against practical skill may indicate that a child has a deep understanding of the “why” of universals, of rules, but is lacking knowledge of when and how to apply those rules to the unique set of circumstances applicable to individualised cases. In short, an understanding of a child’s practical knowledge and technical knowledge taken together, along with an appreciation for the theoretical ‘when’ and ‘what’ of knowledge and learning (the epistemic) provides teachers, parents and students with a more complete understanding of student knowledge.

Implications and Future Research

When working with quantitative data such as this, it cannot be known for certain what in particular may have contributed to the mismatch between standardised assessment and teacher judgement of achievement standard outcomes. Given the subjective, reflective, situated nature of teachers’ work, it would be worth delving deeper and interviewing teachers to determine what their judgements were based upon.

In the Connolly et al. (2012, p. 17) study mentioned earlier it was revealed that some teachers “who were not committed to moderation and the use of standards, or were insufficiently informed about them, preferred to remain neutral in their attitude”. It would be of interest to investigate if the fact that very close to 50% of students in both Years 2 and 3 were awarded a “C” grading in the present study, was in part attributable to a similar lack of shared understandings or commitment from the teachers of the students who participated.

The similarity between the end of Year 2 teacher judgement results and the end of Year 3 teacher judgement results raises another question: Do the majority of children typically develop at a steady pace or do expectations and the desire for consistency play a role? It would be interesting to interview these teachers to see if the grades they awarded were based purely on the assessments they conducted and their knowledge of their students or if the grades awarded by past teachers played a part when planning and reporting. In 1965, Rosenthal and Jacobson (1968) conducted an experiment that sought to explore whether teacher expectation plays a role in student outcomes. The researchers told teachers that one group of students in each year level had greater potential to show high levels of academic growth than another when in actual fact students had been randomly chosen. What Rosenthal and Jacobson came to find was that the students who were expected to show high levels of academic growth showed greater gains in pre and post IQ tests, while the group who were expected to show lower levels of growth showed less; most especially in first, second and fourth grades. The results of that study have since been referred to as the “Pygmalion effect” or “the self-fulfilling prophecy effect”. Though the methods used in the initial study have not been without criticisms, subsequent studies that have sought to examine this same phenomenon have reported similar results (Burgess & Greaves, 2013; de Boer, Bosker, & van der Werf, 2010). In light of this, it would be of interest to investigate whether ‘expectations’ played a role in the present study’s teacher judgement outcomes as well.

Research has also found however that teachers make use of a variety of different sources when making their judgements (Connolly et al., 2012; Wyatt-Smith et al., 2010), sources which may very well include past results or practice and standardised assessment outcomes. It is appreciated that these technicolored, multifaceted profiles play a significant role in the teacher judgement process. Further research comparing teacher judgements of their own students with the variety of sources used to inform the grades they give would shed more light on the findings of the present study.

Conclusion

When working to a continuum, this study has revealed that teachers in the Australian state of Victoria did not echo a similar belief in the spread of their students' abilities as was suggested in those students' standardised assessment results. There have been arguments that have suggested that the standardised assessment may be considered of superior value given its capability for objective, scientific measurement and comparability (Earl, 2005; Klenowski & Wyatt-Smith, 2012). Others have argued that once off standardised assessments are inferior to the longer-term, considered assessment of a teacher who has drawn from a variety of sources when making judgements on student achievement (Harris et al., 2013; Klenowski, 2012). This article has put forward a case for the suggestion that the mismatch between teacher judgement and standardised assessment outcomes may be a result of the fact that the reasoning or knowledge assessed in standardised assessments (the technical) differs to the reasoning assessed by teachers when reporting against a standardised curriculum (the practical). Further to this, it has also been suggested that rather than viewing one as superior to the other, there needs to be an acknowledgement of their complementary relationship.

Addressing the challenge of "student learning", the Victorian Curriculum and Assessment Authority (2014) propose that there needs to be a multidimensional approach which takes into account the "what" (curriculum), the "how well" (assessment) and the "where" (reporting) of student learning (p.7), along with the "how" (pedagogy) which describes "how students will be taught and supported to learn" (p.7). Student learning and knowledge, and reporting, are complex and multifaceted entities. Awarding each dimension an equal value in an era of high stakes testing may go some way towards ensuring that students are being taught in the most productive, informed and supportive of environments.

Acknowledgements

The author would like to thank Prof Brenton Doecke for his mentorship and the Department of Education and Early Childhood (DEECD) and participating schools for their support of this project. The opinions detailed in this article are those of the author and are not necessarily reflective of those held by the DEECD or participating schools.

Maria Nicholas is currently lecturer in Education (Language and Literacy) in the School of Education at Deakin University. She has an interest in literacy development in the early years and the primary/elementary years of schooling. This includes interest in digital literacies, curriculum design and planning, assessment and feedback, and teacher education. Maria's current role follows 13 years of teaching, leadership and consultancy in primary and secondary school settings, as well as a secondment to the position of School Improvement Partner for the Department of Education and Early Childhood Development.

References

- ACARA, Australian Curriculum, Assessment and Reporting Authority 2013a, Guide to understanding 2012 Index of Community Socio-educational Advantage (ICSEA) values.
- ACARA, Australian Curriculum, Assessment and Reporting Authority 2013b, *The Shape of the Australian Curriculum Version 4.0*. Sydney, NSW: Australian Curriculum, Assessment and Reporting Authority Retrieved from http://www.acara.edu.au/verve/_resources/The_Shape_of_the_Australian_Curriculum_v4.pdf.
- Aristotle 350BC/2000, *Metaphysics* (W. D. Ross, Trans.). Raleigh, N.C.: Alex Catalogue.

- Australian Bureau of Statistics 2012, Still on the Move: Reflecting a Nation: Stories from the 2011 Census. cat. no. 2071.0. from <http://abs.gov.au/ausstats/abs@.nsf/Lookup/2071.0main+features702012-2013>
- Australian Curriculum, Assessment and Reporting Authority, ACARA 2013, Results and reports: How to interpret. from <http://www.nap.edu.au/results-and-reports/how-to-interpret/how-to-interpret.html>
- Biesta, Gert 2013, Knowledge, judgement and the curriculum: on the past, present and future of the idea of the Practical. *Journal of Curriculum Studies*, 45(5), 684-696. doi: 10.1080/00220272.2013.798839
- Burgess, Simon & Greaves, Ellen 2013, Test Scores, Subjective Assessment, and Stereotyping of Ethnic Minorities. *Journal of Labor Economics*, 31(3), 535-576.
- Carr, Wilfred 2005, The Role of Theory in the Professional Development of an Educational Theorist. *Pedagogy, Culture and Society*, 13(3), 333-346.
- Connolly, Stephen, Klenowski, Valentina, & Wyatt-Smith, Claire Maree 2012, Moderation and consistency of teacher judgement: teachers' views. *British Educational Research Journal*, 38(4), 593-614.
- Cooksey, Ray W, Freebody, Peter, & Wyatt-Smith, Claire 2007, Assessment as Judgement-in-Context: Analysing how teachers evaluate students' writing 1. *Educational Research and Evaluation*, 13(5), 401-434.
- de Boer, Hester, Bosker, Roel J, & van der Werf, Margaretha PC 2010, Sustainability of teacher expectation bias effects on long-term student performance. *Journal of Educational Psychology*, 102(1), 168.
- DEECD, Department of Education and Early Childhood Development 2013a, English Developmental Continuum P-10. from <http://www.education.vic.gov.au/school/teachers/teachingresources/discipline/english/continuum/pages/default.aspx?Redirect=2>
- DEECD, Department of Education and Early Childhood Development 2013b, Mathematics Developmental Continuum P-10.
- DEECD, Department of Education and Early Childhood Development 2013c, Science Continuum P-10.
- Department of Education and Training 2014, Student Reporting. from <http://www.education.vic.gov.au/school/teachers/support/pages/studentreports.aspx?Redirect=1>
- Doecke, Brenton, Green, Bill, Kostogris, Alex, Reid, Jo-Anne, & Sawyer, Wayne 2007, Knowing Practice in English Teaching? Research Challenges in Representing the Professional Practice of English Teachers. *English Teaching: Practice and Critique*, 6(3), 4-21.
- Earl, Lorna Maxine 2005, *Thinking about purpose in classroom assessment: Assessment for, as and of learning*: ACSA.
- Freebody, Peter, & Wyatt-Smith, Claire 2009, The assessment of literacy: working the zone between 'system' and 'site' validity. *The Journal of Educational Enquiry*, 5(2).

- Harris, Catherine, & Marsh, Colin 2005, Analysing curriculum change: Some reconceptualised approaches. In C. Harris & C. J. Marsh (Eds.), *Curriculum developments in Australia : promising initiatives, impasses and dead-ends* (pp. 15-38). Adelaide, S. Aust.: Openbook.
- Harris, Pauline, Chinnappan, Mohan, Castleton, Geraldine, Carter, Jenni, De Courcy, Michele, & Barnett, Jenny 2013, Impact and Consequence of Australia's National Assessment Program - Literacy and Numeracy (NAPLAN) - Using Research Evidence to Inform Improvement. *TESOL in Context*, 23(1/2), 30-52.
- Kemmis, Stephen 2005, Knowing Practice: Searching for Saliences. *Pedagogy, Culture and Society*, 13(3), 391-426.
- Klenowski, Val 2012, Sustaining Teacher Professionalism in the Context of Standards Referenced Assessment Reform. In A. Luke, A. Woods & K. Weir (Eds.), *Curriculum, Syllabus Design and Equity: A Primer and Model* (pp. 88-102). Hoboken: Routledge.
- Klenowski, Val, & Wyatt-Smith, Claire 2012, The impact of high stakes testing: The Australian story. *Assessment in Education: Principles, Policy & Practice*, 19(1), 65-79.
- Kostogriz, Alex, & Doecke, Brenton 2011, Standards-based accountability: reification, responsibility and the ethical subject. *Teaching Education*, 22(4), 397-412. doi: 10.1080/10476210.2011.587870
- Macken-Horarik, Mary 2011, Building a knowledge structure for English: Reflections on the challenges of coherence, cumulative learning, portability and face validity. *Australian Journal of Education*, 55(3), 197-213.
- Mercurio, Antonio 2005, The discourse of alignment: curriculum and assessment disjunctions at upper secondary level. In C. Harris & C. J. Marsh (Eds.), *Curriculum developments in Australia : promising initiatives, impasses and dead-ends* (pp. 127-144). Adelaide, S. Aust.: Openbook.
- Packer, Martin J, & Goicoechea, Jessie 2000, Sociocultural and constructivist theories of learning: Ontology, not just epistemology. *Educational Psychologist*, 35(4), 227-241.
- Plato 380BC/2013, *Republic* (J. L. Davies & D. J. Vaughan, Trans.). Ware: Wordsworth Editions Ltd.
- Rosenthal, Robert, & Jacobson, Lenore 1968, Pygmalion in the classroom. *The Urban Review*, 3(1), 16-20.
- Sfard, Anna 1998, On two metaphors for learning and the dangers of choosing just one. *Educational researcher*, 27(2), 4-13.
- Snyder, Ilana 2008, *The Literacy Wars: Why teaching children to read and write is a battleground in Australia*. Crows Nest, Australia: Allen and Unwin.
- Sweetman, Edward, Long, Charles R., & Smyth, John 1922, *A history of state education in Victoria: Melbourne* : Published for Education Dept. of Victoria by Critchley Parker, 1922.

- Victorian Curriculum and Assessment Authority 2009, Victorian Essential Learning Standards: Standards and Learning Focus Statements. from <http://pandora.nla.gov.au/pan/129125/20121206-0015/vels.vcaa.vic.edu.au/overview/standards.html>
- Victorian Curriculum and Assessment Authority 2012, Victorian Essential Learning Standards: English. from <http://pandora.nla.gov.au/pan/129125/20121206-0015/vels.vcaa.vic.edu.au/vels/english.html>
- Victorian Curriculum and Assessment Authority 2014, F–10 curriculum planning and reporting guidelines. Melbourne: Victorian Curriculum and Assessment Authority.
- Wyatt-Smith, Claire, Klenowski, Val, & Gunn, Stephanie 2010, The centrality of teachers' judgement practice in assessment: A study of standards in moderation. *Assessment in Education: Principles, policy & practice*, 17(1), 59-75.
- Yates, Lyn, & Collins, Cherry 2010, The absence of knowledge in Australian curriculum reforms. *European Journal of Education*, 45(1), 89-102.