Competing interests? NAPLAN and middle schooling assessment practices

Donna Pendergast and Katharine Swain

Abstract

In Australia, the introduction of the National Assessment Program for Literacy and Numeracy (NAPLAN) in 2008 marked a national shift towards an accountability-driven education agenda. Administered in Years 3, 5, 7 and 9, the tests sit predominantly across the grades commonly known as the ‘middle years’. Both advocates for and critics of NAPLAN note the potential of the test regime to impact on teaching and learning practices, especially as schools and systems strive to reach benchmarks. Coinciding with the introduction of NAPLAN testing, the nation experienced a renewed interest in middle years’ curriculum, pedagogy and assessment practices. This is evident in, for example, the Melbourne Declaration for Schooling (MCEETYA, 2008), which advocates the need for appropriate educational practices in the middle years to ensure optimal learning outcomes for young Australians. Among the practices is a call for authentic and reflective assessment with high expectations, evidenced by higher-order thinking challenges. It is within this context that this study of the 2009 NAPLAN test instruments of the literacy domain for Years 5, 7 and 9 was undertaken. Using Bloom’s Revised Taxonomy (Krathwohl, Bloom, & Masia, 1973) as the analytic framework, each test item was assessed to determine the cognitive complexity of the item. The analysis reveals that less than 10% of questions in the reading tasks afforded the cognitive challenge of higher-order thinking. It is within this context that this study of the 2009 NAPLAN test instruments of the literacy domain for Years 5, 7 and 9 was undertaken. Using Bloom’s Revised Taxonomy (Krathwohl, Bloom, & Masia, 1973) as the analytic framework, each test item was assessed to determine the cognitive complexity of the item. The analysis reveals that less than 10% of questions in the reading tasks afforded the cognitive challenge of higher-order thinking, and hence the interests of middle years’ education and NAPLAN testing might offer some tension in our education system. This paper presents a snapshot in time of a NAPLAN test and raises questions about the degree to which it contributes to the goal of increasing the cognitive challenge of learning and assessment practices in the middle years. It is argued that the NAPLAN test should be understood to be one insight into student achievement, but not the only or full story reflecting student capabilities. NAPLAN results should be employed in this knowledge and not be seen as a panacea for reforming the education system.
Introduction
National Assessment Program Literacy and Numeracy (NAPLAN) testing commenced in Australian schools in 2008 signifying a shift towards an accountability-driven education system agenda (Lingard B., 2010). Reading, writing, language conventions (spelling, grammar and punctuation) and numeracy are tested annually in three year levels belonging to the ‘middle years’, Years 5, 7 and 9, and a fourth in the early years (Year 3). An overarching expectation of NAPLAN testing is the improved accountability of teachers and schools and ultimately, it is expected, improved outcomes for all learners. However, evidence obtained from the Federal Government’s ‘My School’ website (ACARA, 2013) suggests that for some groups of students very little has changed. In Queensland, for example, the results for students in Years 5, 7 and 9 showed little to no improvement from 2008 to 2012 in the domains of reading, spelling, and grammar and punctuation, though there were marginal improvements reported for the 2012 annual assessment process. This raises questions about the diverse factors that impact upon student success and the ways in which the NAPLAN regime (and related practices) support or undermine attempts to respond to the particular needs of middle years’ students.

Coeiding with the NAPLAN agenda, a renewed focus on teaching and learning for students in the middle years is also on the national - and subsequently the state - agenda. Nationally, the Melbourne Declaration on Educational Goals for Young Australians (MCEETYA, 2008) identifies one of its eight interrelated action areas as ‘enhancing middle years’ development’, which aligns with the 10 to 15-year age group of interest in this paper, in particular highlighting the need for educational approaches that are suitable for young adolescents, as articulated in the following quote:

“The middle years are an important period of learning, in which knowledge of fundamental disciplines are developed, yet this is also a time when students are at the greatest risk of disengagement, pedagogy from learning. Student motivation and engagement in these years is critical, and can be influenced by tailoring approaches to teaching with learning activities and learning environments that specifically consider the needs of middle years’ students (MCEETYA, 2008, p. 10).

Drawing on the work of experts in the middle years field such as Bloom (1956), Barratt (1998), Carrington (2002), Chadbourne and Pendergast (2005) and Lingard (2001; Vale, Chilcott, & wires, 2010; 2010) and, well documented authorities in the area of educational assessment such as Wiggins (1989), Earl (2003) Athanassou and Lamprianou (2002; 2005) and Brady and Kennedy (2009), this paper explores the tension between high-stakes testing and middle years’ approaches to assessment, in particular learning and teaching which involves higher-order thinking strategies and authentic and reflective assessment with high expectations, which are identified by the Middle Years of Schooling Association (MYSAs) Position Paper (2008) as necessary elements to engage young adolescents. Significantly challenging for schools is the alignment of curriculum, pedagogy and assessment procedures to meet the needs of students in the middle years. Carrington (2002) argues that teachers should be encouraged to align pedagogy and assessment practices because assessment in the middle years should be viewed as an integral part of curriculum planning and classroom practice. She makes the distinction between assessment that aligns with academic and academic outcomes in the middle years and traditional tests, which are at odds with the middle school principles. Furthermore, Carrington (2002) argues that ‘traditional’ tests are often competitive and do not allow for the students to demonstrate their learning in such areas as problem solving, critical thinking. Bloom (1956), who is long acknowledged as providing a hierarchy of language which, in action, promotes a range of assessment practices and whose theory underpins much of what is known and unknown about the field of assessment research, describes higher-order thinking as involving problem solving and reasoning. He explains that students use a range of cognitive skills including, but not limited to, deductive and inductive reasoning, hypothesising, comparing, classifying and critiquing when functioning in the higher cognitive domain. His taxonomy, more latterly revised, is a developmental continuum ranging from concrete to abstract level of cognition (Kohn, 1999) and it is used in this research as a basis for making judgments about the level of challenge built into NAPLAN tests that were explored. It would be ideal if the NAPLAN tests and the characteristics of effective middle years’ assessment practices were aligned, and certainly not competing or poles apart.

Literature review
High-stakes testing: background and literature
High-stakes testing may be prominent as a 21st century phenomenon but, in fact, it has a long history, initially connected to the challenge of measuring human intelligence. It originated in France in the late 19th century when Binet used crude methodology, which has subsequently been revealed to lack validity, to ascertain the intelligence of students by measuring the circumference of their heads as a means of diagnosing children in need of special assistance classes (Gould, 1981). In 1904 Binet employed a method involving simple tasks to determine those students requiring special education. German psychologist, Stern, developed this technique further in 1912 by dividing the mental age by the chronological age to calculate an intelligence quotient (IQ). Subsequently, attempts to measure intelligence have been linked to both noble and ignoble intentions. For instance, in 1916 Termann recommended IQ as a way of constructing an efficient society, and this was further promoted by Goddard in 1917 as a means of curtailing the reproduction of the ‘feeble minded’ (Gould, 1981). There is no doubt that such tools to measure cognitive ability had high-stakes outcomes for those being assessed, as their life choices were governed by the outcome of such testing regimes. In today’s environment, high-stakes testing is more likely to be connected with opportunity based on economic rationalist consequences for the allocation of and access to resources such as: funds and teachers; membership of classes; and related purposes. In the United States for example, standardised test outcomes have been used to evaluate principals and teaching staff based on their students’ achievement (Popham, 1999). To date there is no such individual teacher assessment mandated by the Australian federal or state governments; however, with the availability of NAPLAN scores and the continued implementation of the National Curriculum, the likelihood of this occurring for Australian teachers and principals is foreseeable. Consistent with previous practices of connecting the outcomes of testing to opportunity, the potential consequences of high-stakes testing regimes becomes a concern.

Critique of high-stakes testing
Criticism of high-stakes testing, such as NAPLAN, takes many forms, some of which will be shared in this paper. It is argued that high-stakes testing does not lead to improved student results. According to the data at the time this study was conducted in 2009, and generally since that time, Queensland’s middle years’ students were reported as being among the worst performing students in the country in the areas of literacy and numeracy (ACARA, 2013). More recent data indicates moderate improvement in 2012 with Queensland moving from sixth to fifth place out of the eight states and territories. A second and related concern is that, even if high-stakes testing regimes improve performance in tests, they do not necessarily improve learning outcomes. Stiggins (2005) asserts that the results are largely misinterpreted; as such, claims of improved standards lack supporting evidence. That is, an increase in test scores is not indicative of an increase in student learning (William, 2008). Harlen and Deakin Crick (2002) concur, arguing that teachers merely drill content and students develop better test taking and memorisation skills and that this does not necessarily lead to them falsely indicate mastery in areas where no mastery actually exists.

A number of researchers advise that ‘test wisdom’ also results in more informed and better guessing in multiple choice tests and is a threat to the validity of test score interpretation (Freire, Bernardi, Edwards, Pedrotti, & Peyton, 2003). They suggest that coaching and practice can help students raise their scores simply by getting used to the types of questions and by practising test style, thereby developing a skill set related to test wisdom rather than a display of cognitive ability. Test wisdom is described by Griffin and Nix (1991, p. 221) as a student’s capacity to use the characteristics of the test and/or the test situation to receive a high score. A concern is that teachers may incorrectly read this test wisdom as an indication of student learning, interpreting that their students have mastery of these skills and move on to the next level. However, their students may experience difficulty, having not achieved a depth of understanding of concepts, and may be overlooked in terms of potential intervention.

Haladyna, Downing and Rodriguez (2002) agree that test practicing may lead to test wisdom, which will affect the consistency of the test results with repeated testing. Testing to teach displays students’ ability to achieve high results in test taking and consequently may result in a false representation of the cognitive level of the child. If these tests were designed to measure cognitive ability they
would include a larger range of cognitive demands and include Bloom’s higher orders of thinking, incorporating opportunities to display analysing, evaluating and creating skills.

According to Griffin and Nie (1991, p. 20), another key issue concerning validity is ‘test anxiety’. They argue that ‘students cannot perform to the best of their ability when they are upset or anxious’, they further submit that a student’s level of anxiety is dependent on the student’s perception of cognitive demands. Kohn (1999) agree, noting that, if anxiety affects test performance, it can be regarded as a source of invalidity that will distort the test scores. From this perspective, the use of the test scores may be problematic. Hence, the interpretation and use of test scores must be carefully considered; item-level information used inappropriately can lead to misunderstanding of student capability (Messick, 1998).

The third matter of concern is the impact that high-stakes testing has on narrowing the curriculum to focus only on those concepts and level of cognitive skill required to achieve improved test results. Research indicates that high-stakes assessment does not necessarily ensure that the curriculum will be followed, but rather become the curriculum as typically a narrower version of the curriculum (Pedulla, Abrams, Madaus, Russell, Ramos, & Jing. 2003). Furthermore, Linn (2000) warns that assessment systems lose much of their dependability and credibility when high stakes are attached to them. Lingard agrees and indicates that high-stakes assessment has led to a much narrower curriculum (Vale, Chilcott, & wires, 2010). Indeed, in high-stakes testing regimes it is common for teachers to adopt surface rote teaching where regurgitation of mere facts is the outcome and the curriculum lacks depth and complex knowledge required for problem solving and decision-making (Shepard, 2000).

The fourth concern is the reduction of teaching time. Black and William (1998) illustrate this effect using the analogy that it does not matter how much time the farmer spends measuring the pig; the pig will not get any fatter if the farmer does not feed it. When teaching time is diverted to developing test skills and siting tests, the time allocated to teaching is reduced. Critics, such as Gunzenhauser (2003), support the concern that this culture of constantly measuring student performance reduces valuable teaching time and that the accountability movement that places inordinate value on test scores to ensure reaching a single benchmark will lead to the practice of ‘teaching to the test’ instead of teachers focusing on areas needing development and even neglecting the child. Brown, Irving and Keegan (2007, p. 136) warn that ‘many negative consequences accrue because national’ assessment has become high stakes’. They assert that, when the nature of the assessment is high stakes, the teacher pedagogy and learning experiences are subverted to mimic more closely the assessment, with the result becoming more significant than the students taking the test.

Validity

Critics argue that middle years’ teachers are wasting valuable teaching time with assessment tasks with questionable validity (Gronlund & Waugh, 2008). Evidence from other countries, such as the United Kingdom and the United States of America, indicates that, when accountability for educational outcomes is measured solely using a moment-in-time, national, full-cohort test, validity and reliability of the assessment data is questionable, such that it cannot be analysed by policy makers in meaningful ways (QSA, 2009). Adrianou and Lampanious (2005, p. 168) explain that there are at least three types of validity:

Content validity – does the assessment match the content and learning outcomes of the subject?

Criterion validity – does the assessment provide evidence of future achievements or performance in other related subjects?

Construct validity – does the assessment really involve the particular behaviours, thought processes or talents that are said to be assessed?

It would appear that there are other validity issues in relation to NAPLAN, including Lather’s (1986) developmental construct of catalytic validity, which addresses issues of altering reality.

In response to a South Australian teacher dismissed for altering Year 7 NAPLAN test answers, the Australian Education Union State Vice-president, Anne Crawford, highlights the intense pressure on students and teachers because of the publication of NAPLAN data on the My School website. She argues that ‘... test has been given a status it simply doesn’t deserve’ and suggests that the general public cannot be assured that teachers altering NAPLAN test results is not a common occurrence (Owen & Edwards, 2010). A spokesperson for the then federal Education Minister, Julia Gillard, argued that teachers changing results is not widespread. However, the ethics of some educators in some Queensland schools is being questioned after Queensland’s NAPLAN security breaches. An article in the state-wide leading newspaper in Queensland, the Courier Mail (Chilcott & Schulz, 2010), reporting cheating allegations against a principal at a Queensland school points to another kind of validity – Professional Validity. Queensland Education Minister at the time, The Honourable Cameron Dick, is quoted in the Courier Mail (Cartwright, 2011) as saying, ‘[W]e don’t want anyone “breaking the rules”.’ ACARA reinforces this statement when clearly stating in the Code of Conduct in the National Assessment Program Literacy and Numeracy Handbook for Principals that breaches of the National Protocols for Test Administration undermines the integrity of the tests (ACARA, 2011, p. 5).

Middle years’ pedagogy

Because NAPLAN testing is conducted in Years 5, 7, 9, and 10, it is a significant feature in the educational landscape that stretches across the middle years, which typically include Years 5 to 10. It is pertinent to understand the theory underpinning middle schooling principles, which are regarded as providing an appropriate education for young adolescents, in particular with respect to assessment practices.

Young people in the middle years are often broadly characterised by the term ‘young adolescent’. Bahr (2005) suggests that there is no defined meaning of ‘adolescence’ globally and curriculum documents refer to a variety of age groups as the ‘middle years’. Education has no approved definition for ‘middle years’; however, according to the Queensland School Curriculum Council, the middle years are considered to be school Years 5 to 10, typically when aged 10 to 15 years. MYSA, the Australia-wide peak body organisation dedicated exclusively to the education, development and growth of young adolescents, in 2008 released the Position Paper on Middle Schooling: People, Practices and Places, which defines middle schooling as ‘an intentional approach to teaching and learning that is responsive and appropriate to the full range of needs, interests and achievements of middle years’ students in formal and informal schooling contexts’ (MYSA, 2008, p. 1). For this discussion ‘young adolescents’ and ‘middle years’ will encompass all students in the 10 to 15 age bracket or the school Years 5 to 10. The MYSA Position Paper specifies three elements necessary for middle schooling:

1. Clear philosophy relevant to the context
2. Comprehensive range of signature practices to engage young adolescents in relevant, meaningful and challenging learning, along with organisational initiatives to facilitate their implementation, such as:
   • higher-order thinking strategies
   • integrated and disciplinary curricula that are negotiated, relevant and challenging
   • heterogeneous and flexible student groupings
   • collaborative teaching
   • small learning communities that provide students with sustained individual attention in a safe and healthy school environment

   emphasis on strong teacher–student relationships through extended contact with a small number of teachers and a consistent student cohort
   • authentic and reflective assessment with high expectations
   • democratic governance and shared leadership
   • parental and community involvement in student learning

3. Evidence-based approach with clearly articulated outcomes, such as:
   • developing current and lifelong learning attributes
   • enhanced academic outcomes
   • creation of a love of learning

Of particular note in this section are the signature practices – higher-order thinking strategies and authentic and reflective assessment with high expectations. A central platform of middle years’ pedagogy is development of higher-order thinking that focuses on problem solving. It entails contexts where the thought processes needed to solve problems and make decisions represent a complex level of thinking, whereby students transform information and ideas so as to understand and discover new meaning (Wheeler & Haertel, 1993). By using skills that involve analysing, classifying, organising, hypothesising and concluding students are able to manipulate information and ideas. It is the role of the teacher to provide opportunities for students to engage in such activities, both in the learning and assessing domains. The demands of this century require students with developed higher-order thinking skills, who are able to demonstrate the ability
to identify the links between diverse concepts. Moreover, they require skills such as planning and organising. Flexible thinking, creative thinking, innovation, problem solving and ability to engage in new disciplines. Students will need to develop and refine these skills so as to achieve in an information-based economy, flooded with technical advancements (Boyd, 2000) and conflicting messages.

Authentic assessment, indicated in the literature as more conducive to the middle years’ philosophy is based upon the premise that assessment should primarily support the needs of learners (Athanasou & Lampaniou, 2005; Brady & Kennedy, 2009; Wiggins, 1989). The word ‘authentic’ is used in much of the literature in this area often with a variety of interpretations (Airasian & Russell, 1998) proposes that middle school assessment should be authentic in that it directly connects to the curriculum experienced by the students, assesses their level of achievement as individuals, and also assists the teacher in designing more effective teaching and learning experiences. Authentic assessment involves explicit links between the curriculum and the assessment so that the process tests what is taught and leads back to better informed teaching and learning (Barratt, 1998). Brady and Kennedy (2009) suggest that authentic assessment is also reflective of the actual learning in the classroom and beyond the classroom and it focuses on student performance and is able to capture the quality of a student’s work. Hence, in terms of implications for NAPLAN, where tests are administered almost exclusively to middle years’ students, it would be ideal if the signifying practices underpinning pedagogy in the middle years – including authentic and reflective assessment and opportunities to demonstrate higher-order thinking strategies – were mirrored in the testing regime, thereby aligning the learning and the assessment approaches.

In recent years in Queensland, the Flying Start White Paper (DET, 2011) recommended and endorsed the plan to move Year 7 into secondary schools and the introduction of a Junior Secondary sector for Years 7 to 9, which incorporates six guiding principles: distinct identity, quality teaching, student well-being, parent and community involvement, leadership, and local decision-making. This is the result of ongoing concerns about the effectiveness of teaching and learning in the middle years, the current transition to Year 8 in the centre of the middle years, and the poor performance – when compared to other states and territories in Australia – of Queensland students in NAPLAN testing (ACARA, 2013). More than a decade ago related concerns were raised as a result of the Queensland School Reform Longitudinal Study (Lingard, et al., 2001), which revealed that Educate Queensland schools were providing their students with supportive environments; however, they provided a relatively low level of intellectual demand, connectedness to the world and recognition of diversity. The evidence from this study also suggests that the middle years remain subject-centric with low levels of student-centred teaching (‘productive pedagogy’) and curriculum integration, with low-stakes assessment regimes (Mills, et al., 2008). The literature from the field emphasises the supportive environment for middle years’ students is essential, but these students also require intellectual challenge to be engaged. Engaged learning is based upon the premise that perspectives on assessment, the what and the why, are plentiful and defines assessment quite simply as “…the process of gathering information to make informed decisions” (Anderson, 2002, p. 11). In contrast, other researchers do not view assessment quite as simplistically.

Contemporary literature distinguishes between four purposes of assessment: diagnostic, formative, summative, and evaluative. In international literature the terms ‘assessment for learning’ (formative) and ‘assessment of learning’ (summative) have gained prominence. Stiggins (2005) suggests that diagnostic and formative, informal assessments sometimes referred to as ‘criterion referenced measures’ or ‘performance based measures’, should be used to inform instruction. Summative, formal or standardised measures should be used for evaluative purposes, to assess overall achievement, to compare a student’s performance with others at their age or grade, or to identify comparable strengths and weaknesses with peers (Stiggins, 2005).

Airasian and Russell suggest that assessment is a process; however, expand on the notion of merely gathering information to instead encompass ‘strategic selection, synthesis and interpretation of information’ or as ‘assessment decision-making’ (Airasian & Russell, 2008, p. 2). Earl (2003) describes assessment as being very complex, dynamic and highly intricate. From this perspective, assessment should not be viewed as a singular entity, but as having a reciprocal relationship with teaching and learning.

Assessment tools involving traditional tests that fail to align with social and academic outcomes in the middle years are at odds with the middle years’ principles (Queensland Government, 2002). Wiggins (1989) argues that assessment is authentic when it directly examines student performance on worthy intellectual tasks and adds that authentic assessment, unlike traditional testing, presents the student with the full array of tasks that mirror the priorities and challenges found in the best instructional activities. It must reflect the real-world context and include situations that students will be confronted with outside the school context. The diverse range of social skills that are so highly valued in today’s business world are best gauged through performance and cannot be measured through multiple choice or short answer tests (Boyd, 2000) – such as those included in NAPLAN. Traditional tests are often competitive and do not allow for demonstration of the variety of learning that has taken place. Barratt (1998) proposes that middle years’ assessment should be authentic in that it involves explicit links between the curriculum and the testing. This method uses a cyclical approach, whereby the process tests what has been taught and leads back to better informed teaching and learning. Chadbourne and Pendergast (2005) concur with Barratt and suggest that middle schools should move towards non-competitive assessment and exhibitions of student work. However, the delivery and implementation of the new national curriculum indicates that these types of assessment are not awarded the same level of credibility as a means of identifying student achievement.

Education Queensland infers that students who only have access to assessment tasks that require low levels of knowledge and cognition are limited to demonstrations of learning at E or D standard. They suggest that the onus is on the teacher to supply tasks and assessment that allows students access to the full range of standards (DETE, pp. 14-19). The Education Framework identifies that questions requiring lower-order knowledge and/or factual recall – Bloom’s lower order of thinking – will offer opportunities for students to achieve a maximum of an E-grade. To achieve an A-grade the assessment task would need to include questions requiring students to demonstrate conceptual knowledge, reflection, reasoning and communication skills in order to compare and contrast effectively higher order of thinking included in Bloom’s continuum. These types of questions require students to use reflection and reasoning tools to move from one representation to another.

If, as Nagel (2010) argues, the adolescent brain does undergo a stage whereby unused synapses are pruned, it is crucially important that higher-order skills such as analysing, evaluating and creating are used regularly so as they are not discarded. Research indicates that adolescents need to be given opportunities to function at the higher levels of cognition (Chadbourne & Pendergast, 2005). Because BRT compartmentalisations the graduating levels of intellectual skills and behaviour, it became a valuable analytical tool to be used for the process of analysing whether the Years 5, 7 and 9 NAPLAN 2009 literacy tests display characteristics that benefit the development of the middle years’ student.

Bloom’s revised taxonomy - higher-order thinking

As noted earlier, Bloom is recognised internationally in education and other fields as an intellectual leader in the field of assessment, having provided a framework that can be generally applied to consider the value and challenge of assessment. Bloom (1956) identified six cognitive levels of intellectual skills and behaviour important to learning: knowledge, comprehension, application, analysis, synthesis, and evaluation. These six levels involved recognition or recall of knowledge and development of intellectual abilities and skills covered in the cognitive domain. In a more recent review of Bloom’s taxonomy, the addition of actions is included, the argument being that learning involves engagement. This newer taxonomy (Table 1) also moves the evaluation stage down a level so that the highest cognitive level of complexity is ‘creating.’ This taxonomy is a much used way of determining the level attained in assessment tasks in Queensland and around the world, and hence is the tool used in this research project.
The learner breaks information down into its component elements. The learner makes decisions based on in-depth reflection, criticism and assessment. The learner creates new ideas and information using what has been previously learned.

Further analysis of the data revealed the remaining questions required only lower-order thinking skills (remember and understand). These questions require students to receive or recite factual information. Knowledge is provided to the students within the text with students primarily engaged in routine lower-order thinking for much of the testing period. Rarely do they go beyond simple reproduction of knowledge. There is an absence of higher-order thinking skills such as evaluate (judging the value of ideas, materials and methods by developing and applying standards and criteria) and create (the development of original ideas and engaging in creative thinking).

The Year 5 reading tasks, although involving variation of genre, sentence structure and vocabulary, are still well-supported by illustrations. Students are expected, when reading a short narrative, to locate directly stated information, connect and interpret ideas, recognize links between text and illustrations, interpret characters’ nature, behaviour and motivation, and identify cause and effect, which are all skills identified by Bloom as requiring lower-order thinking. The expectations of students when reading other genre differ slightly, but the vocabulary used in the minimum standards is mainly that which is associated with the lower levels of BRT. Table 4 reports the language used to explain task expectations for each genre in the 2009 NAPLAN reading tasks for Years 5, 7 and 9 in relation to BRT. The majority of the task descriptions involve skills related to the two lower levels of lower-order thinking, where students are required to perform such skills as remember and understand, with few falling into the higher-order thinking skills area.

Table 1 Bloom’s Revised Taxonomy (adapted from Krathwohl, Bloom, & Masia, 1973; cited in Kohn, 1999)

<table>
<thead>
<tr>
<th>Cognitive levels of complexity</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lower-order thinking</strong></td>
<td></td>
</tr>
<tr>
<td>Remember</td>
<td>The learner is able to recall, restate and remember learned information.</td>
</tr>
<tr>
<td>Understand</td>
<td>The learner grasps the meaning of information by interpreting and translating what has been learned.</td>
</tr>
<tr>
<td>Apply</td>
<td>The learner makes use of information in a context different from the one in which it was learned.</td>
</tr>
<tr>
<td><strong>Higher-order thinking</strong></td>
<td></td>
</tr>
<tr>
<td>Analyse</td>
<td>The learner breaks information down into its component elements.</td>
</tr>
<tr>
<td>Evaluate</td>
<td>The learner makes decisions based on in-depth reflection, criticism and assessment.</td>
</tr>
<tr>
<td>Create</td>
<td>The learner creates new ideas and information using what has been previously learned.</td>
</tr>
</tbody>
</table>

The analysis reveals that five out of a total of the 126 reading comprehension questions (11%) assessed required skills to analyse (Table 3).

Table 2 Distribution of questions according to year level

<table>
<thead>
<tr>
<th>Test</th>
<th>Year 5</th>
<th>Year 7</th>
<th>Year 9</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading comprehension</td>
<td>35</td>
<td>46</td>
<td>45</td>
<td>126</td>
</tr>
<tr>
<td>Language conventions</td>
<td>35</td>
<td>46</td>
<td>45</td>
<td>126</td>
</tr>
</tbody>
</table>

Table 3 Statistical analysis of NAPLAN 2009 reading data (Years 5, 7 and 9)

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of Questions</th>
<th>Lower-Order Thinking</th>
<th>Higher-Order Thinking</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Remember %</td>
<td>Understand %</td>
<td>Apply %</td>
</tr>
<tr>
<td>9</td>
<td>45</td>
<td>40</td>
<td>51.1</td>
</tr>
<tr>
<td>7</td>
<td>46</td>
<td>34.8</td>
<td>63.1</td>
</tr>
<tr>
<td>5</td>
<td>35</td>
<td>37.1</td>
<td>62.9</td>
</tr>
</tbody>
</table>

The writing task was not analysed in this study as it was an open-ended task with little direction that could be assessed in terms of cognitive challenge.

The study
In this study, BRT was used to analyse the literacy component of the 2009 NAPLAN tests administered to Australian middle years’ students (Years 5, 7 and 9). The year was selected for the study as it aligned with the commencement of doctoral candidacy for one of the authors, and serves as a pilot for further work in the doctoral investigation. The items were classified by the authors and researchers using BRT in the relation to the cognitive level of complexity required from students to respond to the questions effectively. One researcher undertook the classification process and the second researcher cross-checked for consistency by randomly selecting a number of items across the data set. The literacy component includes: writing, reading comprehension, and language conventions (spelling, grammar and punctuation) tasks. The writing task was not analysed in this study as it was an open-ended task with little direction that could be assessed in terms of cognitive challenge.

The text analysis identified the term used in the task description, as well as the official expectations in the marking schema used by the assessors. The words/s from within the task and the words highlighting the expectations within the marking guide, were analysed to determine the level of cognitive demand required. The two sections, reading comprehension and language conventions, were assessed by analysing the terms in each question against BRT to determine their respective level of cognitive complexity.

Findings
One hundred and twenty-six questions for each section across Years 5, 7 and 9 were assessed in this study. Table 2 illustrates the distribution between year levels of questions in the reading comprehension and language conventions components of the 2009 NAPLAN tests.

Middle Years of Schooling Association – Volume 13 – Number 1 – May 2013 – www.mysa.org.au
The 2009 NAPLAN language conventions tests were designed to assess the areas of spelling, grammar and punctuation – only the surface accidentals of meaning making. All of the 126 questions were positioned comfortably in the lowest two levels of the BRT. The students were expected to identify and locate common grammatical conventions and recognise punctuation. In order for students to reach minimum standards for the spelling component of the test, at all three year levels, they are required to demonstrate performance of lower-order thinking skills, the lowest level of BRT. The test is a measure of whether the students are able to identify and correct errors in frequently used words; remembering the correct spelling. Results of the NAPLAN Language Conventions Tests have been investigated in some detail in this study: authentic and reflective assessment involving high expectations and higher-order thinking. The need for such practices, tailored for middle years' students, is supported nationally. Yet, using the highly regarded BRT to analyse the NAPLAN test items, NAPLAN provides little opportunity for students to demonstrate higher-order thinking. Furthermore, the genre of the test itself means that it cannot fulfill another of the signifying practices of middle years’ education, which is ‘authentic and reflective assessment with high expectations’ (MYSA, 2008, p. 1). This style of task typically precludes authenticity as understanding of the students’ work within and beyond the classroom requires the development of relationships between the test author/s and the students who are sitting the test, and is often context, site and/or individual specific. Hence, both of the key signifying practices associated with assessment in the middle years were not prominent in the analysed NAPLAN test; therefore, participation in NAPLAN testing in the middle years in 2009 would not have aligned strongly with the principles of middle schooling. Importantly, the test itself is not the only commitment made to the test regime, with many schools allocating considerable time in the weeks preceding the tests to develop test wiseness. Teaching and learning opportunities are likely to mirror the test in order to prepare students adequately for what is anticipated in the tests.

Importantly, NAPLAN does not exist in isolation as an assessment regime in Australian schools, so it is unreasonable to expect it to be an ideal model for middle years’ assessment. However, given that it is used as a benchmarking and reporting mechanism to determine individual, school, state and national levels of literacy and numeracy achievement, it constitutes a high-stakes assessment practice in Australian schools and one that is driving pedagogical practices in the classroom.

This high-stakes, full-cohort, standardised, norm referenced, paper-based test regime has led to an assessment-driven approach to curriculum (Brown, 2008; Chadbourne & Pendergast, 2010).
Australia has followed the lead of other nations, such as that United Kingdom and the United States of America, with the implementation of high-stakes testing. Yet, paradoxically, Finland, which features two major differences to other western countries – flexibility in curriculum delivery and the absence of consequential accountability involving national testing – has consistently demonstrated exceptionally high attainment rates in the international arena (Kupiainen, Hauatan, & Karjalainen, 2009). Perhaps policy direction in assessment in Australia has the opportunity to lead in this direction, by abandoning high-stakes testing as the sole source of information about student achievement, and looking to include a more comprehensive collection of middle years’ assessment practices. There is no doubt that, in current forms, middle years’ education and NAPLAN appear to serve competing interests. All students need to learn how to transform information and accepted ideas so as to understand and discover new meaning in the real world. Having conducted this research, it is recommended that more extensive analysis be conducted over a number of years in order to provide more than a single snapshot in time, which may lead to distortion of the overall test regime, of the cognitive complexities of the NAPLAN test items. Furthermore, we contend that the NAPLAN test outcomes should be understood to be one insight into student achievement, but not the only or full story reflecting student capabilities, and that the use of the results be employed in this knowledge and not be seen as a panacea for reforming the education system.

About the authors:

Prof Donna Pendergast Head and Dean School of Education and Professional Studies Griffith University, Queensland

d.pendergast@griffith.edu.au

Katharine Swain
Doctoral Candidate
School of Education and Professional Studies
Griffith University

References


Willon, D. (2000). What do you know when you know that you know? The meaning of educational assessments. Annual Conference of the International Association for Educational Assessment, Cambridge, IAEA.