Thompson, G. and Harbaugh, A.G. (2012) The effects of NAPLAN: Teachers perceptions of the impact of NAPLAN on pedagogy and curriculum. In: AARE-APERA 2012, 2 - 6 December 2012, University of Sydney, Sydney.

The Effects of NAPLAN: Teacher perceptions of the impact of NAPLAN on pedagogy and curriculum[i]

Greg Thompson and Allen G. Harbaugh
School of Education (Murdoch University, Perth)

Abstract

This paper reports preliminary findings of a survey of in-service teachers in WA and SA conducted in 2012. Participants completed an online survey open to all teachers in WA and SA. The survey ran for three months from April to June 2012. One section of the survey asked teachers to report their perceptions of the impact that NAPLAN has had on the curriculum and pedagogy of their classroom and school.

Two principal research questions were addressed in this preliminary analysis. First, is the socioeconomic drawing area of the school, the State in which they teach, or the school system in which the teacher works significant in perceptions of the impact of NAPLAN on curriculum and pedagogy? Second, are there any interaction effects between gender, socioeconomics status, location and school system on teachers perceptions? Statistical analyses examined one- and two-way MANOVA to assess main effects and interaction effects on teachers' global perceptions. These were followed by a series of exploratory one- and two-way ANOVA of specific survey items to suggest potential sources for differences among teachers from different socioeconomic regions, states and systems.

Teachers report that they are either choosing or being instructed to teach to the test, that this results in less time being spent on other curriculum areas and that these effects contribute in a negative way on the engagement of students. This largely agrees with a body of international research that suggests that high-stakes literacy and numeracy tests often results in unintended consequences such as a narrow curriculum focus (Reid, 2009; Au, 2007), a return to teacher-centred instruction (Barret, 2009; Polesel, Dulfer, & Turnbull, 2012; Barksdale-Ladd & Thomas, 2000) and a decrease in motivation (Ryan & Wesinstein, 2009). Preliminary results from early survey respondents suggests there is a relationship between participant responses to the effect of NAPLAN on curriculum and pedagogy based on the characteristics of which State the teacher taught in, their perceptions of the socioeconomic status of the school and the school system in which they were employed (State, Other non-Government, and Independent).

Introduction

Since 2008 school children in Years 3, 5, 7, and 9[ii] in Australia have sat a series of standardised tests that aim to measure students' basic skills in literacy and numeracy. These tests consist of four domains; Reading, Writing, Language Conventions (spelling, grammar and punctuation) Numeracy (ACARA, 2011). These tests have proved to be a divisive issue for Australian education; some stakeholders have argued that National Assessment Plan – Literacy and Numeracy (NAPLAN) brings a raft of negative consequences to the education experience of young people. Supporters argue that NAPLAN promotes accountability and transparency, measures the basic skills that should be taught,

and the publication of results via the My Schools website equips parents with the information required to make decisions about their children's education.

This paper reports preliminary findings of an online survey of in-service teachers in WA and SA conducted in 2012. This paper utilises a quantitative design and statistical analysis of survey data to examine teacher perception across a large sample. This approach complements much of the emerging quantitative research examining the effects of NAPLAN (Comber, 2012; Wigglesworth, Simpson, & Loakes, 2011). After an initial demographic response schedule, participants were asked their perception of the impacts that NAPLAN has had on their achievement goal orientations, stress levels, self-efficacy to teach and effects on curriculum and pedagogy. These themes were chosen because research literature from the US and UK suggest that teachers perceive significant impacts of high-stakes testing in these areas.

This paper focuses on teacher perceptions of the impact NAPLAN has had on the curriculum and pedagogy of their classroom and school. These perceptions are analysed in relation to demographic information supplied by teachers to see whether these perceptions of effects on curriculum and pedagogy where influenced by their perception of the socioeconomic drawing area of their school, the State in which they worked, their gender or the school system (State, Other non-Government or Independent) in which they taught.

In terms of the overall responses to the impact of NAPLAN on curriculum and pedagogy, teachers in WA and SA largely perceived similar effects to those reported in international studies. The 759 responses tended to suggest that the majority are either choosing or being instructed to teach to the test, that this results in less time being spent on other curriculum areas and that these effects contribute negatively to the engagement of students. This largely agrees with a body of international research that suggests that high-stakes literacy and numeracy tests often results in unintended consequences such as a narrow curriculum focus (Reid, 2009; Au, 2007), a return to teacher-centred instruction (Barret, 2009; Polesel, Dulfer, & Turnbull, 2012; Barksdale-Ladd & Thomas, 2000) and a decrease in motivation (Ryan & Wesinstein, 2009).

The Current State of Knowledge

Current published research into NAPLAN has tended to be focused on theoretical perspectives that link NAPLAN to the intensification of standardised testing in countries like the US and England (Reid, 2009; Lingard, 2010). It is only now that empirical research into NAPLAN is emerging in Australian schools (Klenowski, 2011; Comber, 2012; Ford, 2012). This paper further augments this emerging body of empirical research into NAPLAN, whilst still locating NAPLAN in the broader context of international experiences of high-stakes testing. In particular, this paper will focus specifically on the research that addresses the impacts of that testing on classroom curriculum and pedagogy.

NAPLAN is considered to be a 'high-stakes' standardised test because the results, and subsequent publication of them on the My Schools website which enables comparison between schools, can have significant impacts on parental choice, teacher and principal job security and funding for the school (Polesel, Dulfer, & Turnbull, 2012). NAPLAN carries "serious consequences for students or educators; schools may be judged by their school-wide average; high results bring public praise; while low results may bring public embarrassment" (Lobascher, 2011, p. 10). Lingard (2010) points to the experience in Queensland where low 2008 NAPLAN resulted in widespread, systemic change to the ways that schools, principals and teachers approached the tests as an indication of the high-stakes involved.

Impact on pedagogy

International research suggests that high-stakes testing in schools can potentially have a negative impact on the pedagogy of teachers in countries that use this method of assessment. In many countries

teachers perceive they are required to change their teaching style and strategies (Polesel, Dulfer, & Turnbull, 2012; Au, 2008). Often the pressures to do well manifest as a return of teacher-centred pedagogies that have a negative impact on student creativity and higher-order thinking skills (Lobascher, 2011). One of the unintended consequences of this teacher-centred pedagogy to prepare students for NAPLAN is increased teacher frustration as teachers' professional expertise, knowledge and understanding of each student, teaching methodologies and social support strategies that assist students are marginalised (Klenowski, 2011). This Australian experience is supported by US studies that show teachers who feel pressure to produce students 'up to standards' were more likely to use lecturing styles, directing, and praise/criticism teaching techniques (Barret, 2009; Ryan & Wesinstein, 2009).

Studies in the USA raise concerns about the 'one-size' fits all manner in which these tests and the standards expected create an environment where students who needed extra support or different teaching styles to support their learning needs fail to have those needs met (Ryan & Wesinstein, 2009). Furthermore, many teachers feel that their ability to choose appropriate pedagogy is restricted by their need to address the content for tests rather than individual learning needs (Barret, 2009). Australian studies have found that there is a negative impact on time allocated to struggling students due to the teachers' need to address testing requirements. Often these disengaged students are further disadvantaged as they are ignored as teachers "previous approaches to students with difficulties…used successfully in the past, are temporarily overlooked" so that these students often do not receive the support they need (Comber, 2012, p. 129). Other studies suggest that extra teacher attention and time is given to students who it is felt would very quickly achieve greater results with additional support (Klenowski & Wyatt-Smith, 2012).

Impact on Curriculum

Another impact of high-stakes testing is a squeezed or narrowed curriculum (Reid, 2009). Research suggests that there has been a reduction in subject diversity and offerings and decreased emphasis placed on subjects not measured as a result of standardised testing (Polesel, Dulfer, & Turnbull,

2012). In the USA, the implementation of the No Child Left Behind policy has seen an increasing reduction of specialist class time in schools as more and more class time is devoted to test preparation (Baker, et al., 2010; Phelps, 2007; West, 2012). This reduction of specialist class time to address testing needs is also found in the UK (Collins, Reiss, & Stobart, 2010) and in Australia (Comber, 2012; Polesel, Dulfer, & Turnbull, 2012). Other studies indicate that teachers modified their curriculum to address the tests only (Barret, 2009; Perrault, 2000). This was also found in the UK (Collins, Reiss, & Stobart, 2010), and in Australia (Polesel, Dulfer, & Turnbull, 2012).

In Australia concerns are raised that this narrowing of curriculum is leading to a less diverse and creative classroom environment (Smeed, Spiller, & Kimber, 2009). This concern is also reflected in the experiences of high-stakes testing in the USA (Au, High-Stakes Testing and Curricular Control: A Meta-Synthesis , 2007; Barksdale-Ladd & Thomas, 2000). Research has shown that often the subjects squeezed out of a curriculum are those that support student achievement. For example, there is a positive correlation between student achievement on high-stakes literacy and numeracy testing and engagement with creative subjects (Baker R. , 2012; West, 2012). An Australian study raises concerns that creative, diverse and individualised learning opportunities will disappear from classrooms when curriculum is narrowed to meet external targets (Smeed, Spiller, & Kimber, 2009). The argument follows that with this loss of creativity and individuality due to a narrowed curriculum focus, a loss of higher-order thinking skills results and this loss of higher-order thinking skills will most likely lower student achievement in future tests (Barksdale-Ladd & Thomas, 2000; Polesel, Dulfer, & Turnbull, 2012; Klenowski & Wyatt-Smith, 2012).

A further impact of a narrowed curriculum focus is a fragmentation of knowledge. Studies in the USA, UK and Australia indicate that curriculum planning can lose its long-term coherence, such that themes and scaffolded knowledge are broken down into ‘test-size’ pieces or fragmented knowledge in order to better prepare students for the tests (Barret, 2009; Collins, Reiss, & Stobart, 2010; Polesel, Dulfer, & Turnbull, 2012). In the USA, some schools report making testing practice and teaching to the test part of the daily classroom routine (Barksdale-Ladd & Thomas, 2000). This is also identified in Australian studies (Klenowski & Wyatt-Smith, 2012; Polesel, Dulfer, & Turnbull, 2012). It is

suggested that this emphasis on testing practice and preparation lowers student engagement and long-term extrinsic motivation (Ryan & Wesinstein, 2009).

Impact on Students from Disadvantaged Contexts

High-stakes testing often brings with it the promise that it will improve equity in education because it improves transparency and accountability. The argument seems to be that if you test students for basic skills like literacy and numeracy and publish the scores, teachers and students "will each work harder to achieve better results" (Lobascher, 2011, p. 9). The commonly held assumption is that schools in disadvantaged contexts often underperform on measures of educational equity because the quality of teaching and/or the teaching program is to blame. In the Australian context NAPLAN is often framed by bureaucrats and politicians as a means of improving equity (Gillard, 2008b; Gillard, 2008a; McGaw, 2010). In the Australian context this has meant that NAPLAN needs to be seen in the broader context of an attempt to take "seriously the fact that too many students from 'disadvantaged' backgrounds have for too long been short-changed by the education system" (Reid, 2009, p. 4).

The Australian approach mirrors the rhetoric of high-stakes testing in the US and UK as a mechanism for improving educational equity (Lingard, 2010). However, high-stakes testing may have the opposite effect in that it widens the gap between the highest and lowest achieving students as teachers adopt a 'one-size fits all' approach to pedagogy and curriculum due to the testing and accountability pressures (Ladson-Billings, 1997; McCarty, 2009; Baker, et al., 2010). In the US low-poverty schools are 22 times more likely to be high achieving than are the high-poverty schools. (Harris, 2007). Further, low-poverty and low-minority schools are 89 times more likely to be high achieving (Harris, 2007). Au (2009) characterises this as the "zipcode effect": a phenomena where the socioeconomic characteristics of where students live equates with their achievement on high-stakes tests. Rural children and schools often underperform on high-stakes testing (Jimerson, 2005).

In Australia the lowest achieving students are typically those that come from "disadvantaged backgrounds" (Lange & Meaney, 2011). These include students who identify as Indigenous, have disabilities and/or come from poorer households (Lange & Meaney, 2011). Reid has argued that

where a student lives is a better indicator of their NAPLAN results than where they go to school (Reid, 2009). There is emerging research arguing remote Indigenous students in Australia are particularly disadvantaged because NAPLAN tests often assess culturally assumed knowledge and students may speak a non-standard version of English (Wigglesworth, Simpson, & Loakes, 2011).

Research Questions and Methods

Two principal research questions were addressed in this preliminary analysis. First what are teacher perceptions of the effects on NAPLAN on curriculum and pedagogy? Second, are there any relationships between gender, socioeconomic status, location and school system on teachers' perceptions? Statistical analyses examined one- and two-way MANOVA to assess main effects and interaction effects on teachers' global perceptions. These were followed by a series of exploratory one- and two-way ANOVA of specific survey items to suggest potential sources for differences among teachers from different socioeconomic regions, states and systems.

There were 14 questions that asked participants to evaluate the effects NAPLAN had on their pedagogy and the curriculum of their class and school. 10 of these questions were adapted from the schedule of 20 items that indicate "productive pedagogies" (Lingard, Hayes, & Mills, 2003). The other 4 questions focused specifically on impact on curriculum, impact on pedagogy, and teacher perceptions as to whether NAPLAN had improved literacy and numeracy in their classes. The questions are listed in Table 1.

Table 1

Frequency distribution of responses to survey items measuring perceived impact of NAPLAN on curriculum, planning and classroom activities aggregated across all subgroups.

Survey Item	Frequencies (Percent of responding) SD = 0	D = 1	N = 2	A = 3	SA = 4	*M* (*SD*)
(1) NAPLAN has resulted in me giving lessons that improve higher-order thinking and critical analysis by my students.	140 (18.3%)	265 (34.6%)	201 (26.2%)	149 (19.4%)	12 (1.6%)	1.5 (1.0)
(2) NAPLAN has encouraged me to give lessons that prepare students for the tests.	23 (3.0%)	42 (5.5%)	63 (8.2%)	443 (57.8%)	196 (25.6%)	3.0 (0.9)
(3) NAPLAN promotes classroom conversations that lead to sustained dialogue between students and between teachers and students.	163 (21.3%)	300 (39.1%)	195 (25.4%)	99 (12.9%)	10 (1.3%)	1.3 (1.0)
(4) I feel that student learning in language, grammar and technical vocabulary have improved since the introduction of NAPLAN.	108 (14.1%)	267 (34.8%)	214 (27.9%)	162 (21.1%)	16 (2.1%)	1.6 (1.0)
(5) I feel that student learning in numeracy has improved since the introduction of NAPLAN.	115 (15.0%)	284 (37.0%)	261 (34.0%)	100 (13.0%)	7 (0.9%)	1.5 (0.9)
(6) The need to focus on NAPLAN has forced me to take class time away from other curriculum areas.	22 (2.9%)	69 (9.0%)	84 (11.0%)	315 (41.1%)	277 (36.1%)	3.0 (1.0)
(7) The need to focus on NAPLAN has resulted in less emphasis and teaching time devoted to some curriculum areas in all classes at my school.	15 (2.0%)	79 (10.3%)	128 (16.7%)	313 (40.8%)	232 (30.2%)	2.9 (1.0)
(8) NAPLAN facilitates students making connections from what the tests assess to their real-life contexts.	272 (35.5%)	304 (39.6%)	146 (19.0%)	34 (4.4%)	11 (1.4%)	1.0 (0.9)
(9) NAPLAN does not enable students to focus on identifying and solving intellectual and/or real-world problems.	16 (2.1%)	61 (8.0%)	187 (24.4%)	331 (43.2%)	172 (22.4%)	2.8 (1.0)
(10) NAPLAN has meant that students have control over the pace, directions and outcomes of lessons in my class.	302 (39.4%)	339 (44.2%)	103w (13.4%)	18 (2.3%)	5 (0.7%)	0.8 (0.8)
(11) NAPLAN promotes a socially supportive and positive classroom environment.	310 (40.4%)	306 (39.9%)	125 (16.3%)	22 (2.9%)	4 (0.5%)	0.8 (0.8)
(12) The criteria for judging student performance in NAPLAN are clear and explicit.	94 (12.3%)	216 (28.2%)	187 (24.4%)	246 (32.1%)	24 (3.1%)	1.9 (1.1)
(13) NAPLAN promotes classes where students are engaged and on task.	238 (31.0%)	280 (36.5%)	177 (23.1%)	66 (8.6%)	6 (0.8%)	1.1 (1.0)
(14) NAPLAN facilitates classes where the participation of students of different backgrounds is encouraged.	291 (37.9%)	277 (36.1%)	170 (22.2%)	27 (3.5%)	2 (0.3%)	0.9 (0.9)

Note: Items representing a reversed valence are indicated with italics; for all items, $n = 767$ represents all respondents with no missing data as of 27-May-2012.

Results

Prior to conducting MANOVA omnibus tests to confirm the presence of main effects due to gender, state, school system or SES, the dependent measures were analysed for evidence of common sources of variation. With the 14 curriculum-related survey items[1], principal component analysis suggested that there might be two latent constructs accounting for 48.2% of the observed variance among the survey items. While this preliminary analysis will not be fully explored here, a few comments are noteworthy. First, as seen in Table 2, two items may be minimally related to the emergent constructs (#9 with weak crossloadings) or related to an alternate construct (#12 with weak loadings). Second, the nature of potential latent constructs supports the multivariate analysis as this method allows for an analysis of variation with the presence of multicolinearity. Finally, even though the focus here is on differences in response patterns for each survey item, evidence for latent constructs in fewer dimensions supports further analysis of the source of any observed differences among the individual items.

Table 2

Rotated component matrix for the 14 survey items measuring perceived impact of NAPLAN on curriculum, planning and classroom activities

	Component	
Survey item	1	2
Item 5	.767	
Item 4	.736	
Item 3	.692	
Item 13	.692	
Item 11	.684	
Item 14	.673	
Item 1	.636	
Item 8	.612	
Item 10	.534	
Item 9 (R)	.487	.389
Item 12	.449	
Item 6 (R)		.858
Item 7 (R)		.817
Item 2 (R)		.687

Note: Extraction method was principal component analysis followed by Varimax rotation with Kaiser normalization; loadings below .25 not displayed (largest such loading omitted is less than .18).

[1] For the purpose of the exploratory factor analysis, four items (#s 2, 6, 7 & 9) were reversed coded to maintain a common valence among all survey items.

A MANOVA omnibus test was run on the subset of data respondents without any missing data (n = 767). The break down of sample size within each demographic for the main effects assessed are listed by factor and level in Table 3. Average responses to items grouped by main effects are presented in Table 4. Using Wilk's Lambda test, significant differences were detected for all four of the main effects: by gender, $\Lambda = .949$ with $F(14,747) = 2.88$, $p < .001$; by state, $\Lambda = .954$ with $F(14,747) = 2.57$, $p = .001$; by school system, $\Lambda = .924$ with $F(28,1494) = 2.14$, $p = .001$; and by SES, $\Lambda = .922$ with $F(28,1494) = 2.22$, $p < .001$. With all of the Λ statistics above .92, this suggests that caution must be applied to the interpretation of statistically significant differences. As this is an exploratory analysis attempting to locate potential differences, attention must be paid to effect sizes and potential influence of chance (particularly in light of the sample size and number of hypothesis tests conducted).

Table 3

Sample sizes for each level of the main effect factors.

Factor	Level	Count
Gender	Male	174
	Female	593
State	WA	458
	SA	309
School System	Government	472
	Independent	111
	Other non-Govt	184
SES	Low	67
	Average	656
	High	44

The results of the MANOVA omnibus test suggest that further exploration of the survey items is warranted to locate potential differences among the groups of survey respondents. With the preliminary analysis indicating the possibility of two underlying latent constructs accounting for a fair amount of the variance among the survey item responses, it is reasonable to focus attention on either or both of these item subsets. Due to space limitations in this report, the following discussion will focus on the smaller subset of items. These items are related to perceptions around curriculum effects (whereas the other subset of items relates to perceptions around pedagogy).

Table 4

Average scored responses to survey items measuring perceived impact of NAPLAN on curriculum, planning and classroom activities by main effects.

	Gender		**State**		**School System**			**SES**		
Survey Item	M	F	WA	SA	Gov.	Ind.	Oth.	Low	Avg	High
(1) improve higher-order thinking and critical analysis	1.28	1.58	1.57	1.44	1.49	1.57	1.55	1.66	1.50	1.50
(2) lessons prepare students for tests	2.89	3.00	3.03	2.89	3.10	2.75	2.80	2.94	2.99	2.73
(3) promotes classroom conversations & sustained dialogue	1.15	1.39	1.34	1.34	1.33	1.34	1.36	1.03	1.38	1.20
(4) learning in language, grammar & vocabulary have improved	1.47	1.67	1.72	1.47	1.68	1.67	1.44	1.22	1.66	1.73
(5) learning in numeracy has improved	1.39	1.51	1.54	1.39	1.50	1.59	1.36	1.06	1.51	1.68
(6) class time taken from other curriculum areas	2.90	3.01	3.06	2.87	3.08	2.68	2.92	3.19	2.99	2.59
(7) less time devoted to some curriculum areas in all classes	2.80	2.89	2.98	2.71	2.99	2.57	2.75	3.13	2.86	2.66
(8) connections between tests & real-life contexts	0.98	0.96	0.99	0.93	0.94	1.06	0.98	0.67	1.00	0.86
(9) does not enable students to identify & solve problems	2.69	2.78	2.74	2.78	2.84	2.60	2.65	3.06	2.72	2.89
(10) students control pace, direction & outcomes of lessons	0.79	0.81	0.83	0.77	0.79	0.85	0.82	0.69	0.82	0.75
(11) socially supportive & positive classroom	0.82	0.84	0.86	0.78	0.81	0.80	0.92	0.64	0.85	0.80
(12) clear & explicit criteria for judging student performance	1.82	1.87	1.91	1.77	1.88	1.90	1.76	1.88	1.87	1.64
(13) promotes classes where students are engaged & on task	1.24	1.08	1.20	0.99	1.10	1.26	1.06	0.93	1.14	1.11
(14) encourages participation of students of different backgrounds	1.03	0.89	0.97	0.85	0.88	1.11	0.91	0.66	0.95	0.91

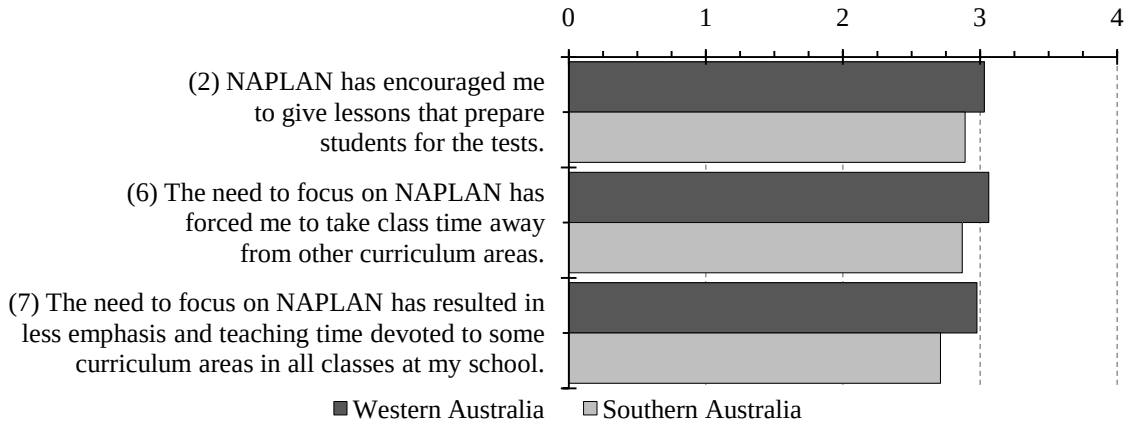
Note: Items representing a reversed valence are indicated with italics; Gov. = Government schools, Ind. = Independent schools, and Oth. = Other non-Government schools; for all items, $n = 767$; standard deviations not individually reported to conserve space, though support of homoscedasticity across items and groupings was evident with values ranging from 0.7 to 1.3 with an average of $SD = 1.0$.

Proceeding from the omnibus test, ANOVAs and t-tests were conducted to locate all items with statistically significant differences between the levels for each factor. Survey items demonstrating statistically significant differences in response patterns for each group are highlighted (all comparisons with $p > 0.05$ are not mentioned).

Gender differences were detected with the omnibus test conducted using all of the survey items. However, when restricted to the four items composing the curriculum subset under consideration here, no significant differences were detected, $\Lambda = .995$ with $F(4,757) = 1.02$, $p = .398$.

State differences were detected for three of the items under consideration (#2, #6 and #7). For the item about preparing students for the test (2nd item), W.A. had an average score of $M = 3.0$ ($SD = 0.9$) and S.A. was 0.1 lower ($M = 2.9$, $SD = 0.9$), $t(765) = 2.10$, $p = .036$. For the item about taking class time from other curriculum areas (6th item), W.A. had an average score of $M = 3.1$ ($SD = 1.1$) and S.A. was 0.2 lower ($M = 2.9$, $SD = 1.0$), $t(765) = 2.52$, $p = .012$. For the item about less time spent on other curriculum areas (7th item), W.A. had an average score of $M = 3.0$ ($SD = 1.0$) and S.A. was 0.3 lower ($M = 2.7$, $SD = 1.0$), $t(765) = 3.57$, $p < .001$.

Figure 1



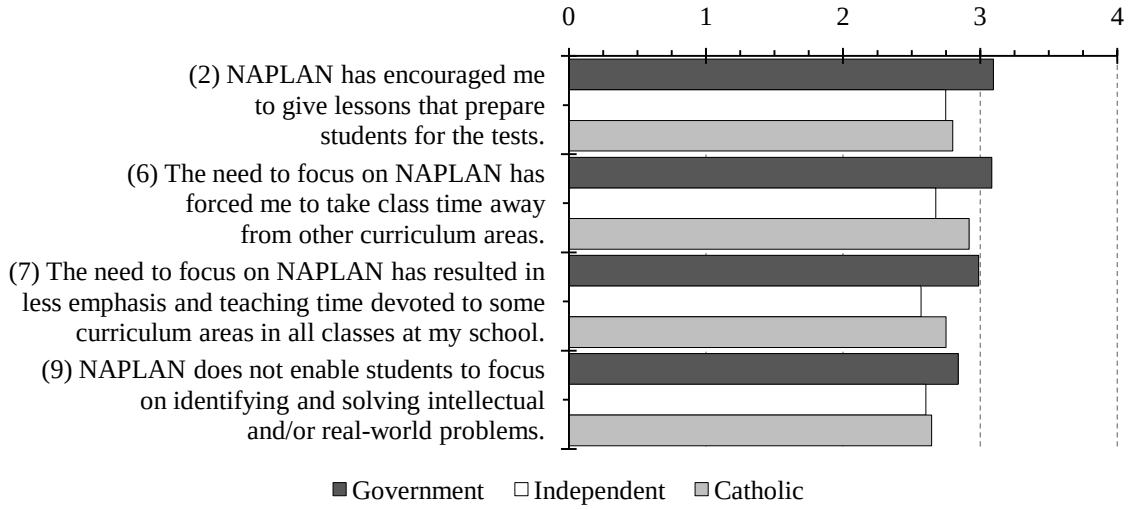
Note: Strongly disagree is anchored at 0 and strongly agree is anchored at 4; W.A., $n = 458$; S.A., $n = 309$.

Comparison of survey item responses for States

Differences among the school systems were examined using 1-way ANOVA tests with 3 levels (government, independent and other non-Government schools). Supplementary analysis revealed that significant differences emerged for government schools compared to either independent schools or Other non-Government schools or both. No differences emerged between independent and Other non-Government schools; a 1-way, 2-level MANOVA on all 14 survey items resulted in $\Lambda = .931$ with $F(14,280) = 1.48$, $p = .117$. As results were comparable between the post hoc analyses for the 3-level and 2-level analyses, it was determined that future comparisons for school systems could be based on government schools versus non-governmental schools for ease of interpretation (though

graphs will display all three categories). School system differences were detected for all four items under consideration (#2, #6, #7 and #9). For the item about preparing students for the test (2nd item), government schools had an average score of $M = 3.1$ ($SD = 0.9$) and other schools were 0.3 lower ($M = 2.8$, $SD = 0.9$), $t(765) = 4.74$, $p < .001$. For the item about taking class time from other curriculum areas (6th item), government schools had an average score of $M = 3.1$ ($SD = 1.0$) and other schools were 0.3 lower ($M = 2.8$, $SD = 1.0$), $t(765) = 3.35$, $p = .001$. For the item about less time spent on other curriculum areas (7th item), government schools had an average score of $M = 3.0$ ($SD = 1.0$) and other schools were 0.3 lower ($M = 2.7$, $SD = 1.0$), $t(765) = 4.11$, $p < .001$. For the item about enabling students to solve problems (9th item), government schools had an average score of $M = 2.8$ ($SD = 1.0$) and other schools were 0.2 lower ($M = 2.6$, $SD = 1.0$), $t(765) = 2.95$, $p = .003$.

Figure 2



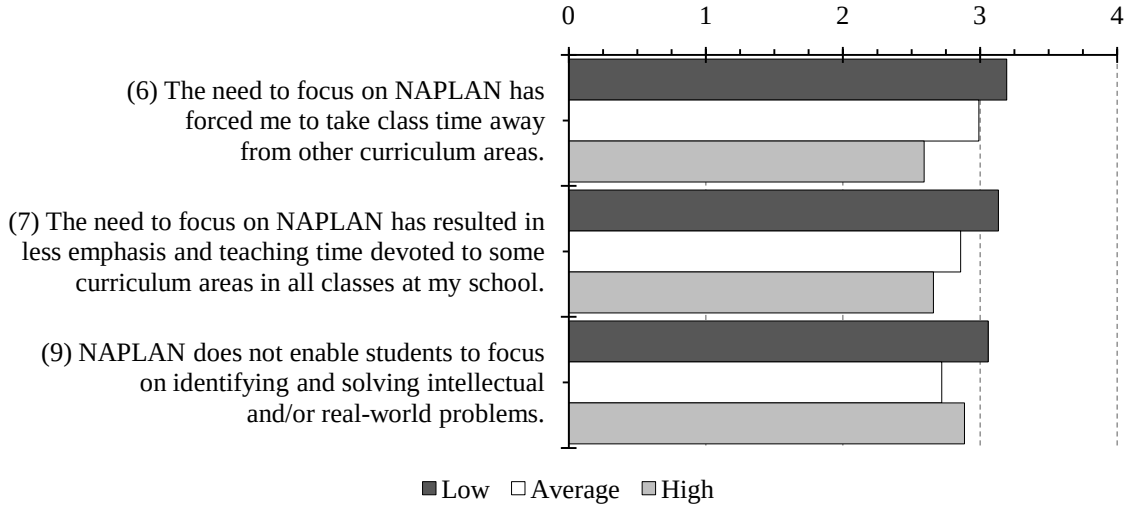
Note: Strongly disagree is anchored at 0 and strongly agree is anchored at 4; government, $n = 472$; independent, $n = 111$; and Catholic, $n = 184$.

Comparison of survey item responses for school systems

SES differences were detected for three of the items under consideration (#6, #7 and #9). For the item about taking class time from other curriculum areas (6th item), the means were significantly different across the three groups with $F(2,764) = 4.53$, $p = .011$ and with the high group differing from both the average and low groups. For the item about less time spent on other curriculum areas (7th item), the

means were significantly different across the three groups with $F(2,764) = 3.25$, $p = .039$ and with the low group differing from both the average and high groups. For the item about enabling students to solve problems (9th item), the means were significantly different across the three groups with $F(2,764) = 4.29$, $p = .014$ and with the low group differing from the average group.

Figure 3



Note: Strongly disagree is anchored at 0 and strongly agree is anchored at 4; low SES, $n = 67$; average SES, $n = 656$; and high SES, $n = 44$.

Comparison of survey item responses for SES

With the four main effects, there would be 6 pairs for 2-way interactions. When examined in both multivariate and univariate analyses, no interaction effects were observed for any of the four items under consideration.

Discussion

Teacher perceptions of the effects of NAPLAN on curriculum and pedagogy as outlined in Table 1 support the findings of international research that high-stakes testing has a number of impacts on curriculum and pedagogy (Au, 2009). Though not fully examined here, there is evidence to suggest that pedagogical responses to the test (as outlined by responses to #3, #8, #10, #11, #14) include adopting a teacher-centred style that has flow on effects of promoting less-inclusive classrooms where

students have less of a voice, less time spent on higher-order thinking skills, less conversation between teachers and students occurs for no appreciable improvements in literacy and numeracy (Au, 2007). In terms of curriculum adjustments for the test (as outlined by #2, #6 and, #7) the findings indicate that teachers are narrowing their curriculum, spending less time on curriculum areas not assessed and that this is having a flow-on effect to the authenticity of their learning as outlined by #9. This supports the international literature that one of the unintended consequences of high-stakes test like NAPLAN is that curriculum is doubly impacted; firstly in that time spent on literacy and numeracy are increased at the expense of other areas, and secondly that "the limited content tested by NAPLAN may result in extra time being spent on basic skills at the expense of critical literacy" (Lobascher, 2011, p. 14). Generally teachers perceive that NAPLAN has required them to a) prepare for the tests, b) change their teaching style to a more teacher-centred approach, c) not improved literacy and numeracy, d) lowered student motivation and engagement, e) created a less inclusive classroom environment for students, particularly those who come from the least advantaged circumstances.

The second half of the results section focused on the relationship between #2, #6, #7 and #9 on the question schedule. This is so that relationships between demographics and teacher perceptions of the effects that NAPLAN has had on their curriculum choices in their classrooms can be measured. This is important because it nuances the perceptions of effects, from macro to micro. There is also evidence of a relationship between #1, #3, #4, #5, #8, #11, #12, #13, and #14, but due to the constraints of this paper this will require another paper to examine in detail.

There was no significant difference across gender for perceived impact on curriculum as measured by #2, #6, #7 and #9. Despite the ratio of responses between male and female teachers mirroring the overall demographics of the Australian teacher workforce, there is no discernable relationship between gender and teacher perceptions of the impact of NAPLAN curriculum (ProductivityCommission, 2012).

However, analysis of the data suggests that there is a relationship between the States in which teachers work (either WA or SA) and their perceptions of the effects of NAPLAN on curriculum in their schools and classrooms. Broadly speaking, while teachers in both States tend to perceive a negative impact of NAPLAN on curriculum, this relationship is higher or more pronounced in WA than SA. It is most likely that this difference is a combination of complex political, procedural, systemic and societal factors being played out in local settings. This requires further investigation to build a hypothesis as to why there is this difference between the States.

There were also significant impacts between the school system and the perceptions of impact on curriculum. Based on the survey responses, while all teachers regardless of school system pointed to a negative impact on the curriculum in their class and school (as evidenced by the literature that narrowing curriculum focus teaching to the test has a negative impact on student learning) teachers in government schools reported NAPLAN having a greater impact on curriculum. There were no significant differences between other non-Government and Independent teacher perceptions. Further research is required to explore the hypothesis that these differences are due to different systemic approaches and emphasis placed on NAPLAN.

There were significant interactions between teacher perceptions of the SES of the school and how they reported changes to the curriculum in their class. This is perhaps the most prominent finding of this research. One of the main aims of NAPLAN testing is to provide data to remove education inequities. However, these responses suggest that the impact on curriculum of NAPLAN is most keenly felt in schools in low socioeconomic drawing areas. If, as research suggests, the broadest curriculum that encourages a range of learning experiences is crucial in lowering the equity gap, then the fact that this is occurring with a greater intensity in those low SES schools may mean that the gap will further grow as a result of NAPLAN. Certainly decades of educational research in the US would suggest that high-stakes tests like NAPLAN actually further disadvantage many students (Au, 2008; Au, 2009; Harris, 2007; Ladson-Billings, 2006; Lange & Meaney, 2011). However, this hypothesis can only be tentatively supported by the survey data, more research is needed to contextualise individual schools with curriculum approaches and NAPLAN results.

Conclusion

This paper has used preliminary data gathered from a survey of in-service teachers in WA and SA that asks their perceptions of the impact of NAPLAN on their school community. In particular it focused one section of the survey that asked teachers their perceptions of the impact of NAPLAN had had on their curriculum and pedagogic choices. This analysis of the preliminary data suggests that teachers perceive NAPLAN having a profound impact on the curriculum and pedagogy in their class and school. Generally teachers perceive that NAPLAN a) has required them to prepare for the tests, b) change their teaching style to a more teacher-centred approach, c) not improved literacy and numeracy, d) lowered student motivation and engagement, e) created a less inclusive classroom environment for students, particularly those who come from the least advantaged circumstances.

The second half of the paper examined whether there were any interaction effects between the gender of the teachers, the system in which they taught, the socioeconomic status of the school and teachers perceptions of the impacts of NAPLAN on curriculum and pedagogy. The analysis suggests that there is no discernable interaction between teacher gender and perception, but that there are significant interactions between the system in which a teacher works and their perception of impact and the socioeconomic status of the school and teacher perception of impact. While teachers surveyed, regardless of system and school SES reported a prominent impact of NAPLAN on curriculum, the interactions were most pronounced in State schools and low SES schools respectively. Whilst this is only preliminary data and more research is required to explain why these patterns may be emerging, this is concerning in a testing regime that aims to promote equity of educational outcomes.

References

ACARA. (2011). *NAPLAN*. Retrieved 2012 йил 29th-May from NAPLAN: http://www.nap.edu.au/NAPLAN/About_each_domain/Writing/index.html

Au, W. (2008). Devising inequality: a Bernsteinian analysis of high-stakes testing and social reproduction in education. *British Journal of Sociology of Education, 29* (6), 639-651.

Au, W. (2007). High-Stakes Testing and Curricular Control: A Meta-Synthesis . *Educational Researcher*, 258-267.

Au, W. (2009). *Unequal By Design: High-Stakes Testing and the Standardization of Inequality.* New York: Routledge.

Baker, E., Barton, P., Darling-Hammond, L., Haertel, E., Ladd, H., Linn, R., et al. (2010). *Problems with the Use of Student Test Scores to Evaluate Teachers.* Washington: Economic Policy Institute.

Baker, R. (2012). The Effects of High-Stakes Testing Policy on Arts Education. *Arts Education Policy Review, 113* (1), 17-25.

Barksdale-Ladd, M. A., & Thomas, K. F. (2000). What's at Stake in High-Stakes testing: Teachers and Parents Speak Out. *Journal of Teacher Education, 51* (5), 384-397.

Barret, B. (2009). No Child Left Behind and the assault on teachers' professional practices and identities. *Teaching and Teacher Education, 25* (8), 1018-1025.

Collins, S., Reiss, M., & Stobart, G. (2010). What happens when high-stakes testing stops? Teachers' perceptions of the impact of compulsory national testing in science of 11-year-olds in England and its abolition in Wales. *Assessment in Education: Principles, Policy & Practice, 17* (3), 273-286.

Comber, B. (2012). Mandated literacy assessment and the reorganisation of teachers' work: federal policy, local effects. *Critical Studies in Education, 53* (2), 119-136.

Ford, M. (2012). Achievement gaps in Australia: what NAPLAN reveals about education inequality in Australia. *Race Ethnicity and Education*, 1-23.

Gillard, J. (2008b йил 27-March). *Education revolution: COAG agrees to school reform agenda.* Retrieved 2010 йил 17-June from Media Centre: http://mediacentre.dewr.gov.au

Gillard, J. (2008a йил 10-September). *NAPLAN, OECD Report, teacher quality, age pension.* Retrieved 2011 йил 29th-April from Hon Julia Gillard MP: http://mediacentre.dewr.gov.au/mediacentre/Gillard/Releases/NAPLANOECDReportteacherqualityagepension.htm

Harris, D. (2007). High-Flying Schools, Student Disadvantage, and the Logic of NCLB. *American Journal of Education, 113* (3), 367-394.

Jimerson, L. (2005). Placism in NCLB—How Rural Children are Left Behind. *Equity & Excellence in Education, 38* (3), 211-219.

Klenowski, V. (2011). Assessment for learning in the accountability era: Queensland, Australia. *Studies in Educational Evaluation, 37*, 78-83.

Klenowski, V., & Wyatt-Smith, C. (2012). The impact of high stakes testing: the Australian story. *Assessment in Education: Principles, Policy & Practice, 19* (1), 65-79.

Ladson-Billings, G. (2006). From the Achievement Gap to the Education Debt: Understanding Achievement in US Schools. *Educational Researcher, 35* (7), 3-12.

Ladson-Billings, G. (1997). It Doesn't Add Up: African American Students' Mathematics Achievement. *Journal for Research in Mathematics Education, 28* (6), 697-708.

Lange, T., & Meaney, T. (2011). Becoming disadvantaged: Public discourse around national testing. *European Society for Research in Mathematic Education.* Poland.

Lingard. (2010). Policy borrowing, policy learning: testing times in Australian schooling. *Critical Studies in Education, 51* (2), 129-145.

Lingard, B., Hayes, D., & Mills, M. (2003). Teachers and productive pedagogies: Contextualising, conceptualising, utilising. *Pedagogy, Culture & Society, 11* (3), 399-423.

Lobascher, S. (2011). What are the Potential Impacts of High-stakes Testing on Literacy Education in Australia? *Literacy Learning in the Middle Years, 19* (2), 9-19.

McCarty, T. (2009). The impact of high-stakes accountability policies on Native American learners: evidence from research. *Teaching Education, 20* (1), 7-29.

McGaw, B. (2010 йил Tuesday 24th-August). Driving education reform with curriculum reform and accountability. *Banksia Lecture* . Perth: Murdoch University.

Perrault, G. (2000). The classroom impact of high-stress testing. *Education, 120* (4), 705-710.

Phelps, P. (2007). *Standardized Testing.* New York: Lang Publishing.

Polesel, J., Dulfer, N., & Turnbull, M. (2012). *The Experience of Education: The impacts of high stakes testing on school students and their families.* Sydney: Whitlam Insititute.

ProductivityCommission. (2012). *Schools Workforce: Productivity Commission Research Report.* Canberra: Australian Government.

Reid, A. (2009). Is this a revolution? A critical analysis of the Rudd government's national education agenda. *Curriculum Perspectives, 29* (3), 1-13.

Ryan, R., & Wesinstein, N. (2009). Undermining Quality teaching and learning: A Self Determination Theory Perspective on High-Stakes Testing. *Theory and Research in Education, 7* (2), 224-233.

Smeed, J., Spiller, K., & Kimber, M. (2009). Issues for principals in high-stakes testing. *Principal Matters, 81,* 32-34.

West, C. (2012). Teaching Music in an era of high-stakes testing and budget reductions. *Arts Education Policy Review, 113* (2), 75-79.

Wigglesworth, G., Simpson, J., & Loakes, D. (2011). NAPLAN Language Assessments for Indigenous students in Remote Communities: Issues and Problems. *Australian Review of Applied Linguistics, 34* (3), Online.

[i] This research has been made possible by a grant from the ARC

[ii] The data from NAPLAN began to be published online via the My Schools website in 2010.