GOR08534

On the status and quality of Physical Education as school subject: An empirical study of classroom and assessment practice

Jennifer M Gore, James G Ladwig, Wendy Amosa and Tom Griffiths
The University of Newcastle, Australia

As a profession, teaching is prone to status concerns and, within teaching, physical educationists are particularly given to pondering status issues… we also know that there is a subject hierarchy but there is little comparative data on this. What one suspects is that physical education is universally lower rather than higher in the pecking order of school subjects.

(Hoyle, 1986, cited in Sparkes and Templin, 1992, p.124)

Status has long been an issue for Physical Education (PE) as it has struggled for recognition in relation to other school subjects. PE's questionable status relates primarily to perceptions of what counts as legitimate knowledge worthy of a place in the school curriculum. Perceived as "practical" and "non-cognitive" and hence inferior to other school subjects, PE has consistently been denigrated by some policy makers, academics, teachers and parents (Kirk & Tinning, 1990).

During the past couple of decades there have been major shifts in what is taught in schools in the name of Physical Education. Its scientific bases, its links with biological and medical sciences, its invocation in the context of discourses of increasing obesity and lifestyle related illnesses, have all contributed to the reshaping of the subject. In school curricula, these shifts are manifest in its linking with other studies such that we now have 'Health and Physical Education' or where we live, in New South Wales, Australia, 'Personal Development, Health and Physical Education' as school subjects. In its various contemporary and more complex subject guises, Physical Education is now an examinable subject (Green, 2000) at the highest levels in many countries around the world, including Australia and the UK.

Despite these changes in the construction of PE as school subject and its shift from marginal status toward greater acceptance as having a legitimate role in schooling, questions remain about its place in school. The subject is still associated with play rather than work (Apple, 1979) and PE teachers are sometimes maligned, if only jokingly, by some of their colleagues, and by the popular media, as not 'real' teachers. Beginning teachers report the relatively low status as a significant influence on their early experiences (Smyth, 1995) and levels of dissatisfaction (Macdonald, 1995). Moreover, a recent international worldwide survey of the state and status of physical education in schools reported issues of legal status and actual implementation, restricted or decreasing curriculum time allocation, subject status and attitudes of headteachers, other teachers and parents, inadequacies in financial, material and human resources and teacher preparation, curriculum trends, as well as scepticism about the subject's future as major conditions confronting physical education (Marshall & Hardman, 2000).

While at risk of contributing further to physical educationists' preoccupation with status, in this paper we address the issue by providing empirical evidence of the quality and substance of PE teaching relative to other school subjects. In so doing, we craft our arguments about the relative worth of physical education in the manner of the courtroom, rather than the pulpit (Ladwig, 1996). Specifically, we explore differences and similarities between Physical Education and

other subjects as well as variations within Physical Education in terms of conceptions of curriculum, treatment of knowledge, and approaches to pedagogy and assessment. Drawing on a large body of quantitative and qualitative data, collected in secondary (Grades 7-10) classes, our analyses identify potential misconceptions and misunderstandings of physical education, offering both cause for celebration and reason to pause for those aligned with the subject and some potentially surprising insights for those who prefer to view the subject from without (or on high).

Research design

The cross-sectional analyses presented in this paper are drawn from a large-scale longitudinal study of the relationships among teacher professional learning, the quality of pedagogy, and the quality of outcomes for more than 2000 students as they progressed through four years of schooling (2004-2007). The study, titled 'Systemic Implications of Pedagogy and Achievement in NSW Public Schools'[1] (SIPA) included the collection of extensive data within the school subjects of English, Maths, Science, Human Society and Its Environment (HSIE, or as it in known in other contexts, Social Studies) and Physical Development, Health and Physical Education (PDHPE, the current NSW version of Physical Education). English, Maths and HSIE data were collected across Primary (Elementary) and Secondary grades while Science and PDHPE data were collected primarily in secondary schools.

Data collected for the entire study take the form of assessment tasks (n = 523 tasks), the associated class sets of student work produced by these tasks (n = 21521 pieces of work), classroom observations (n = 661), and interviews with teachers (n = 612). Assessment tasks and classroom observations were coded using instruments designed to measure the quality of pedagogy. For this purpose, we used the NSW Quality Teaching instruments (NSW DET, 2003; 2005) which judge levels of Intellectual Quality, Quality Learning Environment and Significance (with six elements in each of these dimensions) (see Table 1 and Appendix 1) using a 5 point scale (see, as an example, the coding scales for Deep Knowledge and Explicit Quality Criteria in Appendix 2). The NSW Quality Teaching model draws on the significant work of Newmann and Associates (1996) on Authentic Pedagogy, as well as our earlier work on Productive Pedagogy (Education Queensland, 2001) which made use of other elements of classroom and assessment practice that have been linked through empirical research to improved learning outcomes for students across the spectrum of social backgrounds. Student work was coded using a modified form of the Authentic Achievement instrument (see Newmann & Associates, 1996) for levels of Problematic Knowledge, Analysis, Depth of Understanding and Elaborated Communication using a 4 point scale (example provided in Appendix 3), since this body of research shows a strong correlation between Authentic Achievement and achievement on conventional standardised tests. Interviews were coded in relation to key research questions of the study, with reference to the subject affiliation and schooling level (Primary/Secondary) of the teachers. With all of these data we have controlled for school-level variables such as the socio-economic status of the student population and the proportion of indigenous students at each school. In comparisons between subjects, analysis of variance was used to determine levels of significance and effect sizes were estimated using Cohen's d and omega squared (ω^2**)**.

[1] See other SIPA papers for details of research design and research questions. These include: Amosa, Ladwig, Griffiths and Gore (2007); Griffiths, Gore, and Ladwig (2006), Gore, Ladwig, Griffiths and Amosa (2007); Ladwig, Smith, Gore, Amosa and Griffiths (2007).

Table 1 The Dimensions and Elements of the Quality Teaching Model

Intellectual quality	Quality learning environment	Significance
Deep knowledge	Explicit quality criteria	Background knowledge
Deep understanding	Engagement*	Cultural knowledge
Problematic knowledge	High expectations	Knowledge integration
Higher-order thinking	Social support*	Inclusivity*
Metalanguage	Students' self-regulation*	Connectedness
Substantive communication	Student direction	Narrative

Note: *Marked elements do not pertain to the coding of assessment practice.

The system context in which this research was conducted is worth noting. The largest state of Australia, New South Wales, governs its 2200 public schools (providing for 760,000 students) with a single state authority, the NSW Department of Education and Training (NSW DET). The NSW DET commissioned the development of a model of pedagogy that could underpin a State-wide initiative to improve the quality of teaching in public schools throughout NSW. The "Quality Teaching" model is being implemented to varying degrees in schools (Gore and Ladwig, 2006) while those schools participating in this study at least expressed interest in being part of the study. Furthermore, the fact that two subject faculties signed up in each school meant there was some enthusiasm among the teachers in those faculties and not just among school leaders. The instruments used for data collection were known to, and often used by, the teachers whose classes we observed and assessment tasks we analysed. While the NSW DET has not mandated implementation of Quality Teaching, the support provided to teachers both centrally and at the school level has highlighted Quality Teaching as a rigorous attempt at system-wide pedagogical reform across all school subject areas.

In our focus on PDHPE in this paper, data were drawn from the five secondary schools in the study that had identified PDHPE as an area of focus (schools were asked to nominate two subject areas). These data pertain to students who were in Years 7-10 during the years 2004 to 2007. Observational (or classroom practice) data are available for 42 PE lessons while task or assessment data relate to 73 PE tasks. We also draw on interviews with 36 PDHPE teachers. The preliminary questions addressed in our analysis for this paper were:

- What kinds of tasks were set by teachers in PDHPE?
- What kinds of teaching were observed in PDHPE ?
- What quality was measured in observations and tasks for PDHPE and how does the quality compare with the total sample?

We address these questions before returning to considerations of the status of PDHPE.

Assessment tasks in PDHPE

The assessment tasks coded for Quality Teaching for all subjects in this study were submitted by teachers as tasks that produced the best work by their students during a particular time period, typically around three to five months. Naturally, teachers' judgements and subject specific differences in the nature of tasks used by teachers impacted on the work submitted and limit generalisations that can be drawn about the overall quality of assessment tasks in particular subjects. On the basis of the work submitted, however, we can report that a majority of the PDHPE tasks were summative in design and substantial in scope, designed to require students to engage with whole units of work they had been studying rather than smaller topics or problems. Tasks varied in content and requirements of students but consistently involved issues of personal and social significance and engaged students in tasks that required deep understanding through

consulting a range of sources and drawing their own conclusions. This characterisation of the tasks is consistent with the perspective on learning that Kirk and O'Flaherty (2003) argue must accompany approaches to authentic assessment. In one of the schools we studied, there had been a concerted effort, under the direction of a new Head Teacher for PDHPE, to rewrite and refine assessment tasks:

> We've been rewriting things furiously and creating and rewriting assessment tasks and making sure that the assessment work that we are doing (and the class work we are doing) is more contemporary, modernized, relevant in today's society. We have updated all of the knowledge we are giving the kids and the current affairs… current statistics and making sure that we assess them based on more contemporary frameworks and incorporating a bit more literacy research and problem solving stuff into their assessment tasks… All of our assessment tasks have been passed around other members of the faculty to cross check things. (825011)

This account of the work done by PE teachers on the quality of assessment tasks is well supported by the tasks themselves that rated particularly highly in this school (on Quality Teaching measures).

The following list of task topics is indicative of the kind of work required of students in PDHPE in our sample:

- Making a difference – letter to council etc
- Relationships and me – Scenario analysis
- Life long habits – Article for a youth magazine
- Behaviours and consequences of health – assignment
- The changing world of me – assignment
- The nature of adolescence – presentation
- Safe living campaign – contextual analysis of effectiveness
- Backyard football – design a game

Teachers we interviewed spoke of the time and energy that both they and their students needed to put in to the assessment side of the subject PDHPE. From one teacher's perspective

> English teachers … would not believe all I did last holidays was mark. Five part assignments and when we get a few classes in each year… if you don't take a fair bit home each week then you fall behind, which would probably be a joke to PE teachers a long time ago. (825003)

This comment acknowledges the common perception of PE as non-academic not only among teachers of other subjects but also among PE teachers of perhaps not so long ago. These accounts of energy going into assessment in PE are indicative of the changing nature of the subject. In the following section we move beyond perceptions to analysis of the quality of PE tasks in comparison with tasks in other subjects.

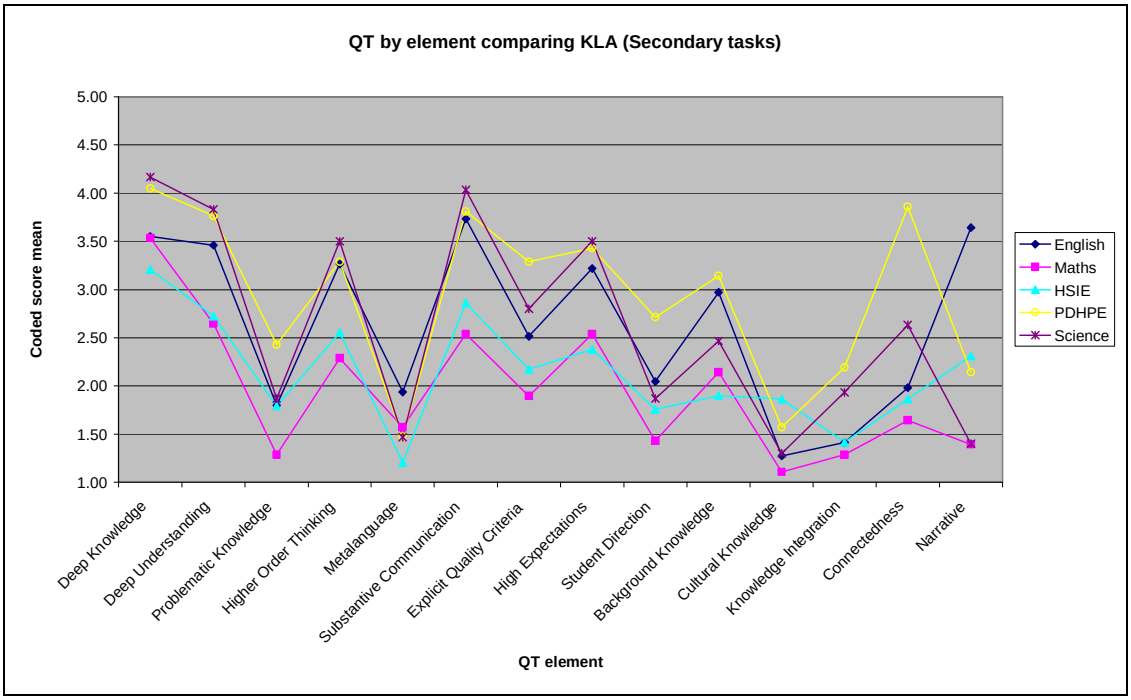
Assessment quality in PE relative to other subjects

In the analysis of the assessment tasks across all five subjects, PE was superior (as measured by the Quality Teaching instrument). That is, when examining the performance of each subject, by dimension of Quality Teaching, higher scores were found for the Quality Learning Environment (QLE) and Significance (SIG) dimensions than in any of the other subjects while PE scored equal highest for the Intellectual Quality (IQ) of its assessment tasks (see Table 2).

Table 2 Assessment Task by subject, QT dimension and QT total (2004-2007)

		QT dimension							
		IQ		QLE		SIG		QT total	
	n	mean	sd	mean	sd	mean	sd	mean	Sd
English	229	18.34	3.77	7.88	2.18	11.38	2.40	37.61	6.49
Maths	60	15.17	4.61	6.08	2.20	7.95	2.88	29.20	7.74
HSIE	52	15.37	3.61	6.69	1.96	10.08	2.46	32.13	6.35
PDHPE	**73**	**18.33**	**3.61**	**9.11**	**1.78**	**13.66**	**2.62**	**41.10**	**6.83**
Science	64	19.47	3.53	8.58	2.63	10.46	2.66	38.69	7.64

Figure 1 illustrates the mean scores by element of the model and demonstrates that higher scores were obtained in PE than in any of the other subjects for the following elements of the model: Problematic knowledge (PK), Explicit Quality Criteria (EQC), Student Direction (SD), Background Knowledge (BK), Knowledge Integration (KI), and Connectedness (C). Furthermore, among the five subjects PE scored second highest for the elements Deep Knowledge (DK), Deep Understanding (DU), Higher Order Thinking (HOT), Substantive Communication (SC), High Expectations (HE), and Cultural knowledge (CK). That is, the only elements where PE tasks were not of the highest or second highest quality among the 5 subjects were Metalanguage and Narrative. Both line graphs and column graphs are used to illustrate these findings, with the column graph being most appropriate for the representation of these data but the line graph easier to read in terms of subject differences.



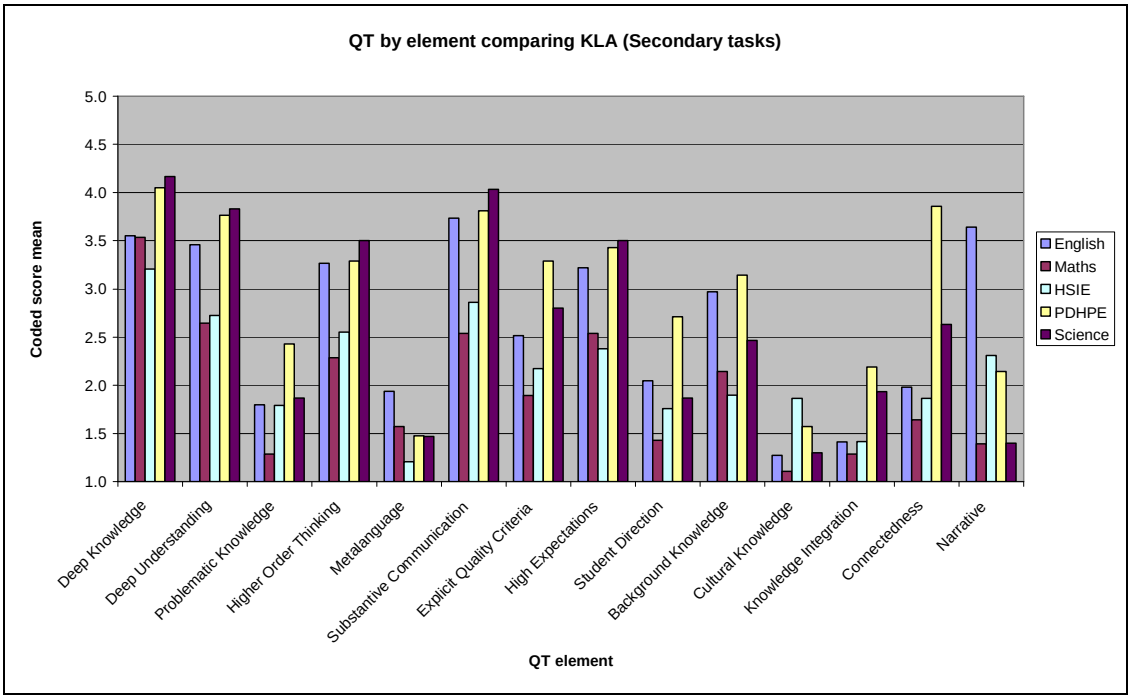


Figure 1. Assessment Task by subject and QT element

Additional analyses were conducted using Cohen's d and omega squared (ω^2) to explore effect sizes and the strength of association between variables (Table 3). Cohen's d enables determinations as to whether statistically significant differences are differences of practical concern. A Cohen's d of 0.2 is considered small, 0.5 moderate, and 0.8 large (Cohen, 1977). Note that most of the Cohen's d reported in Table 3 are at least moderate (cells shaded pale green) if not large (cells highlighted in bright green), with some very large effect sizes obtained. Omega squared (ω^2) is used to provide a measure of the strength of the association between independent and dependent variables. When interpreting ω^2, 0.01 is a small association, 0.06 is a medium association, and 0.14 or larger is a large association (Kirk, R.E., 2006). Note that most ω^2 results show medium (cells shaded pale blue) to large associations (cells shaded bright blue). Table 3 depicts the results of these statistical tests, and reports those elements of the model for which the difference between PDHPE and another subject held up to effect size scrutiny. For example, PDHPE scored higher than HSIE on Deep Knowledge ($p<0.001$, $d= 0.54$, $\omega^2 = 0.06$). In those few instances where another subject scored better than PDHPE the results for that subject are depicted on the table in italics. For example, Science scored higher than PDHPE on Deep Knowledge ($p<0.001$, $d = 0.70$ and $\omega^2 =0.10$). Where PDHPE scored better than another subject, that subject is included in the table using regular font. Subjects where the difference between means was not statistically significant are not listed.

Table 3. Assessment Task by subject with a significant difference between means

Elements	KLA	n	mean	sd	se	Sig (2-tailed)	Cohen's d	ω^2
Deep Knowledge	**PDHPE**	**73**	**3.81**	**0.74**	**0.09**			
	HSIE	52	3.38	0.84	0.12	0.00	0.54	0.06
	Science	*64*	*4.31*	*0.71*	*0.09*	*0.00*	*0.70*	0.10
Deep Understanding	**PDHPE**	**73**	**3.60**	**0.85**	**0.10**			

	Science	*64*	*3.97*	*0.94*	*0.12*	*0.02*	*0.41*	0.03
	HSIE	52	2.92	0.84	0.12	0.00	0.81	0.13
	Math	60	2.90	0.93	0.12	0.00	0.79	0.13
Higher Order Thinking	**PDHPE**	**73**	**3.22**	**0.85**	**0.10**			
	HSIE	52	2.69	0.85	0.12	0.00	0.62	0.08
	Math	60	2.55	0.93	0.12	0.00	0.75	0.12
	Science	*64*	*3.72*	*0.86*	*0.11*	*0.00*	*0.58*	0.07
Problematic Knowledge	**PDHPE**	**73**	**2.41**	**1.27**	**0.15**			
	Math	60	1.40	0.74	0.10	0.00	0.97	0.18
Substantive Communication	**PDHPE**	**73**	**3.78**	**0.69**	**0.08**			
	HSIE	52	2.96	0.84	0.12	0.00	1.07	0.22
	Math	60	2.90	1.19	0.15	0.00	0.91	0.17
	Science	*64*	*4.11*	*0.72*	*0.09*	*0.01*	*0.47*	0.04
Intellectual Quality	**PDHPE**	**73**	**18.33**	**3.61**	**0.42**			
	HSIE	52	15.37	3.61	0.50	0.00	0.82	0.13
	Math	60	15.17	4.61	0.60	0.00	0.76	0.12
Explicit Quality Criteria	**PDHPE**	**73**	**3.21**	**0.88**	**0.10**			
	HSIE	52	2.42	1.04	0.14	0.00	0.81	0.14
	Math	60	1.95	1.23	0.16	0.00	1.18	0.26
High Expectations	**PDHPE**	**73**	**3.19**	**0.86**	**0.10**			
	HSIE	52	2.60	0.82	0.11	0.00	0.71	0.10
	Math	60	2.75	0.84	0.11	0.00	0.52	0.06
	Science	*64*	*3.73*	*1.01*	*0.13*	*0.00*	*0.58*	0.07
Student Direction	**PDHPE**	**73**	**2.71**	**0.96**	**0.11**			
	HSIE	52	1.67	0.86	0.12	0.00	1.14	0.23
	Math	60	1.38	0.94	0.12	0.00	1.40	0.32
	Science	64	1.84	0.96	0.12	0.00	0.90	0.16
Quality Learning Environment	**PDHPE**	**73**	**9.11**	**1.78**	**0.21**			
	HSIE	52	6.69	1.96	0.27	0.00	1.29	0.29
	Math	60	6.08	2.20	0.28	0.00	1.51	0.36
Background Knowledge	**PDHPE**	**73**	**3.27**	**1.35**	**0.16**			
	HSIE	52	1.92	1.06	0.15	0.00	1.11	0.22
	Math	60	2.28	1.04	0.13	0.00	0.82	0.13
Cultural Knowledge	**PDHPE**	**73**	**1.67**	**0.82**	**0.10**			
	HSIE	*52*	*2.15*	*1.04*	*0.14*	*0.01*	*0.52*	0.06
	Math	60	1.15	0.66	0.09	0.00	0.72	0.10
Knowledge Integration	**PDHPE**	**73**	**2.33**	**0.69**	**0.08**			
	English	229	1.45	0.79	0.05	0.00	1.19	0.19
	HSIE	52	1.37	0.56	0.08	0.00	1.53	0.35
	Math	60	1.33	0.73	0.09	0.00	1.40	0.33
	Science	64	2.03	0.82	0.10	0.02	0.39	0.03
Connectedness	**PDHPE**	**73**	**4.01**	**0.54**	**0.06**			
	English	229	2.09	1.11	0.07	0.00	2.21	0.40
	HSIE	52	1.88	0.68	0.09	0.00	3.48	0.75
	Math	60	1.68	0.89	0.12	0.00	3.16	0.72
	Science	64	3.00	1.11	0.14	0.00	1.16	0.25
Narrative	**PDHPE**	**73**	**2.37**	**1.26**	**0.15**			
	English	*229*	*3.57*	*1.40*	*0.09*	*0.00*	*0.90*	*0.12*
	Math	60	1.50	0.79	0.10	0.00	0.82	0.13
	Science	64	1.45	0.82	0.10	0.00	0.86	0.15
Significance	**PDHPE**	**73**	**13.66**	**2.62**	**0.31**			
	English	229	11.38	2.40	0.16	0.00	0.91	0.13
	HSIE	52	10.08	2.46	0.34	0.00	1.41	0.32

	Math	60	7.95	2.88	0.37	0.00	2.08	0.52
	Science	64	10.64	2.66	0.33	0.00	1.14	0.24
Quality Teaching	**PDHPE**	**73**	**41.10**	**6.83**	**0.80**			
	English	229	37.61	6.49	0.43	0.00	0.52	0.05
	HSIE	52	32.13	6.35	0.88	0.00	1.36	0.30
	Math	60	29.20	7.74	1.00	0.00	1.63	0.40

These further analyses confirm the superiority of PDHPE tasks in terms of Intellectual Quality and Quality Learning Environment relative to Maths and HSIE and in terms of Significance relative to all other subject areas. Note that there was not a single instance where another subject was superior to PDHPE at the dimension level. Overall, aggregated Quality Teaching scores found PDHPE was superior to Maths and HSIE (large effect size) and English (moderate effect size), and no instance where another subject was superior.

Interpretation

Views that Physical Education is not intellectually challenging or academically rigorous or of high quality are summarily dispelled by these results. There are several possible explanations for the results that will be explored more fully in subsequent analyses. One explanation is that the intellectually demanding tasks set by these teachers are indicative of qualities and capacities of the teachers themselves who, in Australia, unlike some other countries, often enter universities with entrance scores higher than students in high status areas. For instance, at the University of Newcastle, entrance scores for the PE teaching degree (Bachelor of Teaching/Bachelor of Health and Physical Education) between 2004 and 2006 were higher than for Bachelor of Engineering (Electrical) or Bachelor of Engineering (Chemical) and up to x points higher than for secondary teaching awards in …. other subject areas.

A second explanation relates to the nature of the PE syllabus for Years 7-10 which is organised around key concepts and deep understanding, the writers having made use of the Quality Teaching model in its development, and thus closely aligned with the principles of Quality Teaching (Board of Studies NSW, 2006). In NSW, teachers adhere fairly closely to syllabus documents tending to view them as more prescriptive than indicative of what they might do, at least in terms of content to be taught. Thus, the PE syllabus itself might be more closely aligned than syllabuses in other subjects with the principles of Quality Teaching.

A third and related explanation for the high quality tasks in PDHPE relates to the subject matter being assessed. In particular, high levels of Significance are not surprising given the relation of topics covered in the tasks to students' lives and hence the close alignment of these tasks with this dimension of the QT model. This explanation will be tested through a more detailed content analysis of the tasks in a subsequent paper.

A final explanation we are exploring relates to the richer or more comprehensive nature of the assessment tasks given in PE which tended to be project-based, designed to be completed over several lessons or weeks, and which required students to access information from various sources in order to complete the tasks rather then rely on what was "in their heads".

It is worth emphasising that it was not only in the Significance dimension that the PE tasks scored well – some critics might be quick to dismiss this result as one to be expected. The Quality Learning Environment, as measured by the extent to which tasks conveyed high expectations for students to engage in challenging work, provided explicit criteria for what would count as good

work, and gave students significant influence over some aspects of their work, was also higher in PE than in Maths and HSIE as was the Intellectual Quality relative to these two subjects.

In examining individual tasks, we also saw quality across the various dimensions. That is, in an analysis of the five highest scoring PE tasks for each dimension, there was considerable overlap; one task scored in the top five for all three dimensions, another was in the top 5 for Intellectual Quality and Quality Learning Environment, and a third task was in the top 5 for both Quality Learning Environment and Significance. This overlap is indicative of the comprehensiveness and potential utility of the QT model for the design and refinement of assessment tasks. But what of classroom practice?

Classroom practice in PDHPE

While the relative quality of assessment tasks might not be all that surprising to any readers familiar with what Green (2000) calls the 'academicisation' of the subject, we anticipated different findings in the quality of PE lessons relative to other subjects. The practical lessons observed centred on teams games, including cricket, volleyball, softball, touch football, netball and the slightly less traditional Walla Rugby (a modified version of Rugby Union), and dodgeball. The non team-game practical lessons we observed were lessons in gymnastics, table tennis, fitness testing, and coordination (concluding with a speed ball match). The theory lessons observed centred on water safety, smoking, drug use, risk taking behaviour versus safe living, sexually transmitted infections, and hygiene. Less traditional content included lessons on health issues in Third World countries, health products and services, and community and school roles in physical activity.

Most noteworthy was the fact that the lessons we observed were dominated by fairly traditional approaches to teaching physical education and health as separate subjects largely organised around traditional content. This observation stands in contrast to new PE discourse and syllabus representations of the subject. We draw attention to these data to illustrate that some of the rhetoric associated with curriculum reform in PDHPE was not manifest in the lessons we observed, (which is not to say it is not happening in NSW or elsewhere). While we found clear evidence of teaching that related to the four strands of the PDHPE 7-10 Syllabus – self and relationships, movement skill and performance, individual and community health, and lifelong physical activity – some of the stated "implications for teaching and learning in PDHPE" (p.15) were much less apparent. For instance, we saw little evidence of the following syllabus imperatives:

> PDHPE issues are best dealt with in the context of a comprehensive approach to the learning area. The integration of related outcomes and content reinforces the interrelationship that exists between health and physical activity issues and discourages the teaching of these concepts in isolation.
>
> The combination of integrated programs and student-centred learning approaches ensures that PDHPE issues are addressed in contexts that give them meaning and purpose. It empowers students to acquire deeper understanding that supports more effective development of skills. (Board of Studies NSW, 2006, p.9)

In our interviews with teachers, there were references to integrating aspects of the subject – "what happens is we do fitness testing in practical and in theory at the same time so we try and integrate what they're learning in practical as well as what they're learning in theory" (794008) – or using student centred approaches in their teaching, but the overwhelming evidence gained from hours of observing lessons was much more conventional and separate. Hence, our evidence is

consistent with Tinning's (1996) observation from more than a decade ago: "having a national curriculum statement which provides a framework for the connection between health and physical education is one thing, changing teachers' practice is another" (p.9).

Our interviews with PDHPE teachers help to explain what we observed. One teacher, who originally trained in other subjects but was asked to teach PE in a private school, and was subsequently granted permission to teach PDHPE in the state system, stated "I've been a gym instructor for 10 years so all of the theory side of the PE courses I'd already covered in that [work] and I played a heap of sports as a young guy at a fairly high level so that takes up the practical side" (825011). While we would not suggest that his experience or views were typical of PDHPE teachers, such views certainly contribute to the reproduction of approaches to the subject as being primarily about fitness and sport. Another teacher, in talking about the balance in time allocated to theory and practical classes in the junior secondary years commented: "You're just aiming not to put them off early" (825015). While open to multiple interpretations, this statement, to some extent, reproduces a view of PE as moving negatively toward the worst of other (academic) subjects and indicates both teacher and student preferences for practical rather than theoretical representations of PE.

Another teacher said:

> I believe that subjects were initially developed in order to recognise needs in our society and where kids are moving to and I think of the traditional subjects such as Maths, Science, English as areas providing academic rigour and development of those specific skills that will allow them to learn and educate and be self learners. With the PDHPE, I don't think that was the idea of it in the first instance and I think that our focus area has been spread to such a large degree that now we're being required to look into those fields. Whereas, *I think we should be doing more practical.* I also think there should also be more practical hands-on subjects where we're probably addressing the theoretical aspects in other subjects as well" (825010, emphasis added).

These views help explain the apparent disparity between the official curriculum and the curriculum in use and even the disparity between teachers' talk about the subject and what we observed.

It is important to add that the treatment of PE that we observed as being separate from health or personal development, as being primarily about sports, and as a 'practical subject' does not mean it necessarily involved traditional or outdated pedagogy. Indeed, as one teacher boldly claimed

> seeing … a teacher from another KLA … who has the same teaching methodology you know, 'open up your text books, do question two, four, five, part a through to e', that's fairly static and I don't think that my faculty and I will say it now, I don't think that any PDHPE faculty in New South Wales will be teaching that way, but I could see that maybe other faculty areas could well be.

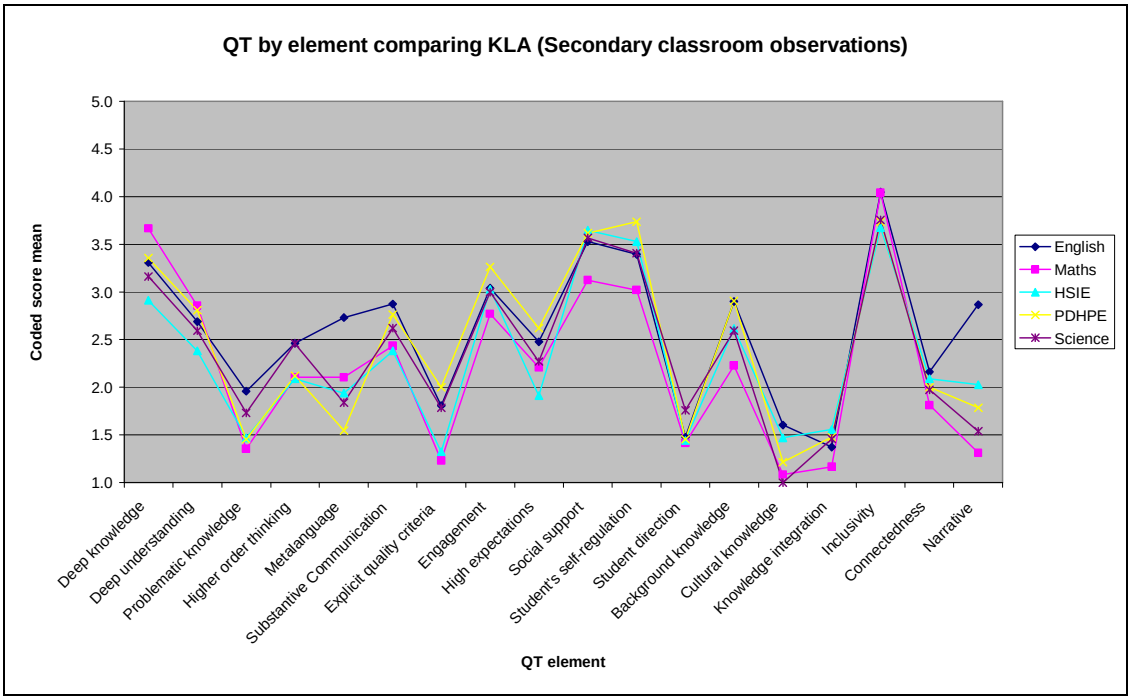
In the next section we consider the extent to which this teacher's claims for the quality of pedagogy in PE hold up to comparison with the other subjects in our study.

Classroom practice quality in PE relative to other subjects

Unlike the assessment tasks, where scores were consistently high across the three dimensions of Quality Teaching, scores for classroom practice in PE were among the lowest of the subjects for Intellectual Quality, were the highest of all subjects in Quality Learning Environment, and third highest among the set of subjects for Significance (see Table 4). Figure 2 depicts the differences among elements and show Engagement and Students' Self Regulation to be the only elements where PDHPE scored highest of all subjects.

Table 4. Classroom observation by subject, QT dimension and QT total (2004-2007)

		Observations QT dimension							
		IQ		QLE		SIG		QT total	
	n	mean	sd	mean	sd	mean	sd	mean	sd
English	143	16.01	4.36	15.70	4.08	14.93	3.66	46.64	10.57
Maths	48	14.52	3.42	13.77	4.02	11.65	2.93	39.94	8.88
HSIE	34	13.18	4.30	14.88	3.62	13.44	3.71	41.50	10.08
PDHPE	42	14.05	3.35	16.69	3.95	13.12	3.42	43.86	8.48
Science	37	14.41	4.57	15.78	4.33	12.32	3.01	42.51	10.72



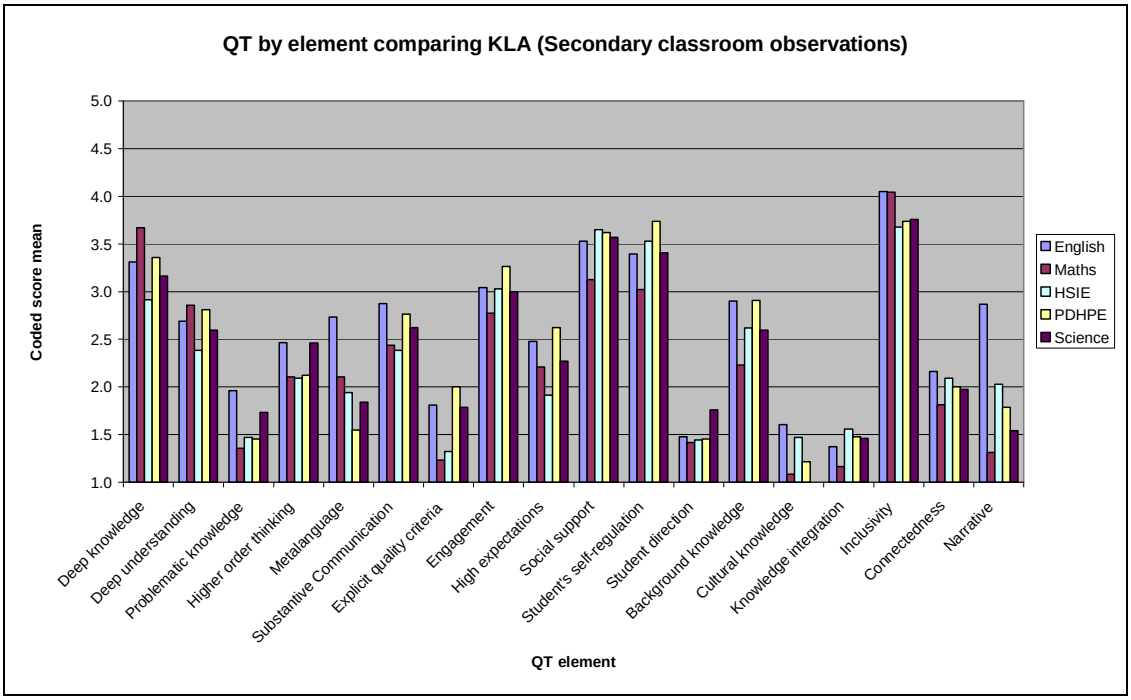


Figure 2. Classroom observations by subject and QT element

Effect sizes are reported in Table 5 and show fewer differences between PDHPE and other subjects than were found for assessment tasks. The only dimension for which differences between subjects are of note is the Quality Learning Environment dimension whereby PDHPE scored better than Maths and HSIE. Note that, as for tasks, there is no dimension for which another subject scored better, using these tests, than PDHPE.

Table 5. Classroom Observation by subject with a significant difference between means

Elements	KLA	n	mean	sd	se	Sig (2-tailed)	Cohen's d	ω^2
Deep Understanding	**PDHPE**	**42**	**2.81**	**0.77**	**0.12**			
	HSIE	34	2.38	0.85	0.15	0.03	0.52	-0.01
Problematic Knowledge	**PDHPE**	**42**	**1.45**	**0.67**	**0.10**			
	English	*143*	*1.95*	*0.95*	*0.08*	*0.00*	*0.61*	*0.05*
Metalanguage	**PDHPE**	**42**	**1.55**	**0.77**	**0.12**			
	English	*143*	*2.73*	*1.19*	*0.10*	*0.00*	*1.17*	*0.16*
	HSIE	*34*	*1.94*	*0.89*	*0.15*	*0.05*	*0.47*	*0.09*
	Math	*48*	*2.10*	*0.86*	*0.12*	*0.00*	*0.68*	*0.04*
Explicit Quality Criteria	**PDHPE**	**42**	**2.00**	**1.08**	**0.17**			
	HSIE	34	1.32	0.53	0.09	0.00	0.79	0.16
	Math	48	1.23	0.56	0.08	0.00	0.90	0.12
High Expectations	**PDHPE**	**42**	**2.62**	**1.06**	**0.16**			
	HSIE	34	1.91	0.90	0.15	0.00	0.72	0.03
	Math	48	2.21	0.92	0.13	0.05	0.41	0.10
Social Support	**PDHPE**	**42**	**3.62**	**0.94**	**0.14**			
	Maths	48	3.13	1.16	0.17	0.03	0.47	-0.01
Students Self Regulation	**PDHPE**	**42**	**3.74**	**0.94**	**0.14**			
	English	143	3.38	1.03	0.09	0.04	0.36	0.02
	Maths	48	3.02	1.12	0.16	0.00	0.69	0.00
Engagement	**PDHPE**	**42**	**3.26**	**0.91**	**0.14**			

Elements	KLA	n	mean	sd	se	Sig (2-tailed)	Cohen's d	ω^2
	Maths	48	2.77	1.06	0.15	0.02	0.50	0.00
Quality Learning Environment	**PDHPE**	**42**	**16.69**	**3.95**	**0.61**			
	HSIE	34	14.88	3.62	0.62	0.04	0.48	0.11
	Math	48	13.77	4.02	0.58	0.00	0.73	0.04
Background Knowledge	**PDHPE**	**42**	**2.90**	**1.05**	**0.16**			
	Maths	48	2.23	0.86	0.12	0.00	0.70	0.01
Cultural Knowledge	**PDHPE**	**42**	**1.21**	**0.52**	**0.08**			
	Science	37	1.00	0.00	0.00	0.01	0.58	0.06

Possible explanations for these findings include – (1) the lower Intellectual Quality of practical lessons versus theory lessons; (2) Significance scores that are affected by the differences between practical and theory lessons; (3) students' continuing pleasure in PE and hence high engagement, self- regulation, social support and so on, especially in a society where sport is so popular; and (4) the organisational and teaching skills of PE teachers that might account for the high scores in Quality Learning Environment. We do not have much direct evidence for Explanation 4, although one of the teachers interviewed referred to ways in which PE teachers (in his school at least) work together that might impact on the quality of teaching practice:

> The nature of the PE staff is that we are very often team teaching classes. So whereas when I taught Maths full time for a year, no other teacher walked into my Maths room, not the principal, not the deputy, not another teacher, it would've been considered rude, or if I'd walked into an experienced Maths teacher's class and said "can I just come and watch your class for 10 minutes?", they'd say, "why?" and they'd say, "no." But in PE every PE teacher would be spending at least 3 or 4 lessons a year with every other PE teacher and with teachers that are on the same year and the same classes. It might be up to 10 to 15 lessons, so all of us know how all of the others teach and we all get ideas from each other. … we get awesome opportunities to team teach and say, "I would have done that differently", or "I like the way you do that", so we're already doing that sort of stuff. (825011)

Explanations 1 and 2 for the classroom practice scores, and for which we have extensive data, address the differences between practical and theory lessons. In the next section of this paper, we consider these differences with reference to the Quality Teaching scores for practical and theory lessons and the interview data gathered from teachers in the study.

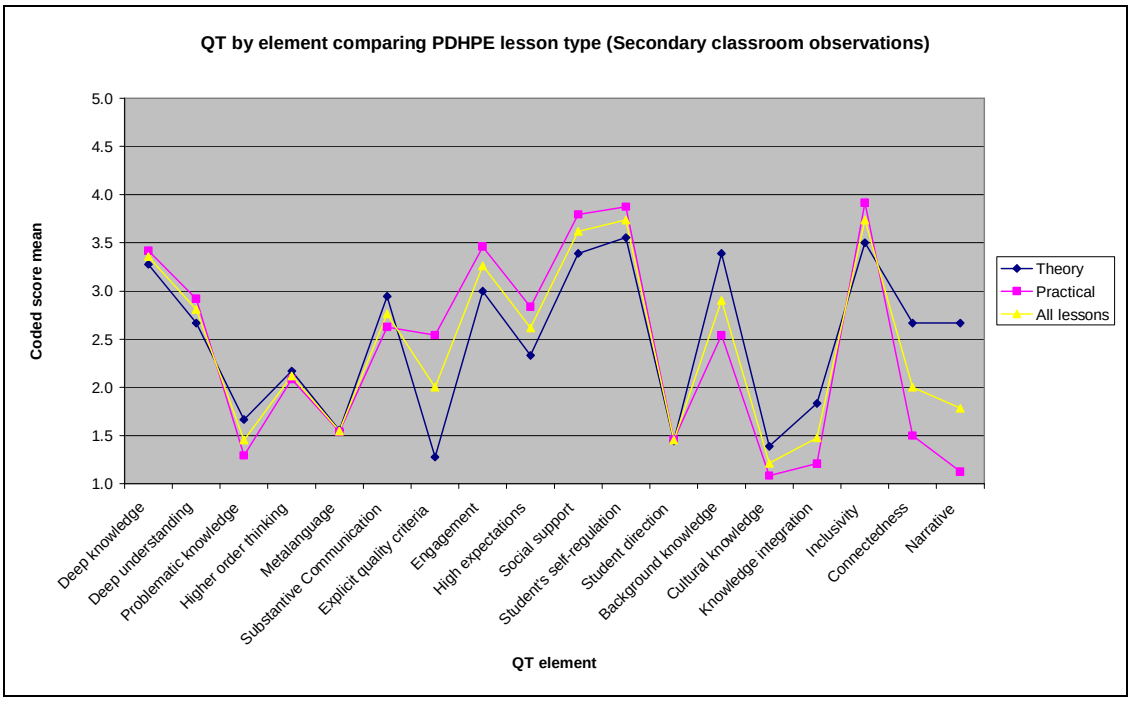
Theory versus practical lessons

'Classroom practice' as a category for this study is complicated in PDHPE by the apparent variation between lessons taught in classrooms and those taught in places like the gym or on the oval. This distinction is commonly referred to by teachers, students, and academics as the distinction between 'practical' lessons and 'theory' lessons. Of the 42 PDHPE lessons for which we have observational data (18 theory lessons and 24 practical lessons) using the QT coding instrument, the four highest scoring lessons were all classroom or 'theory' lessons. Topics for those lessons included:

- two separate lessons introducing a new unit on turning challenges into opportunities and focusing on diet;
- eating and perceptions of health;
- a health lesson on exercise and diet; and
- a hygiene lesson in a girls' school based on a unit of work titled 'managing menstruation.'

The next three highest scoring lessons were all practical lessons in which the focus was: learning skills in cricket; an introduction to the game of volleyball; and, learning the rules and playing a modified version of Rugby Union. The eighth highest scoring lesson was again theory and focussed on health products and services. Hence, in our analysis of the top 20% of lessons, there was a mix of theory and practical classes. Similarly, in the bottom 20%, there was a mix of theory and practical classes with exactly half of the eight lowest scoring tasks being theory lessons and the other half practical. This simple analysis suggests that theory and practical classes might not be as different, in terms of the quality of teaching and intellectual demands of the work, as might be anticipated.

However, in comparing these 42 theory and practical lessons using ANOVA, some statistically significant differences were found between the two lesson types (see Figure 3). Before reporting these results, comparisons that yielded no statistically significant differences are worth higlighting. In particular, despite common perceptions of practical lessons as not as intellectually demanding as theory lessons, our analyses show no statistically significant differences between practical and theory lessons for any of the elements of Intellectual Quality or for the dimension overall (see Table 6). Problematic Knowledge, Higher Order Thinking, and Substantive Communication were slightly lower in practical lessons than they were in theory lessons, but not significantly so. Hence, in this analysis, even practical classes appear to be addressing important concepts and contributing to the development of deep understanding, thus providing further evidence of PE's worthiness among subjects.



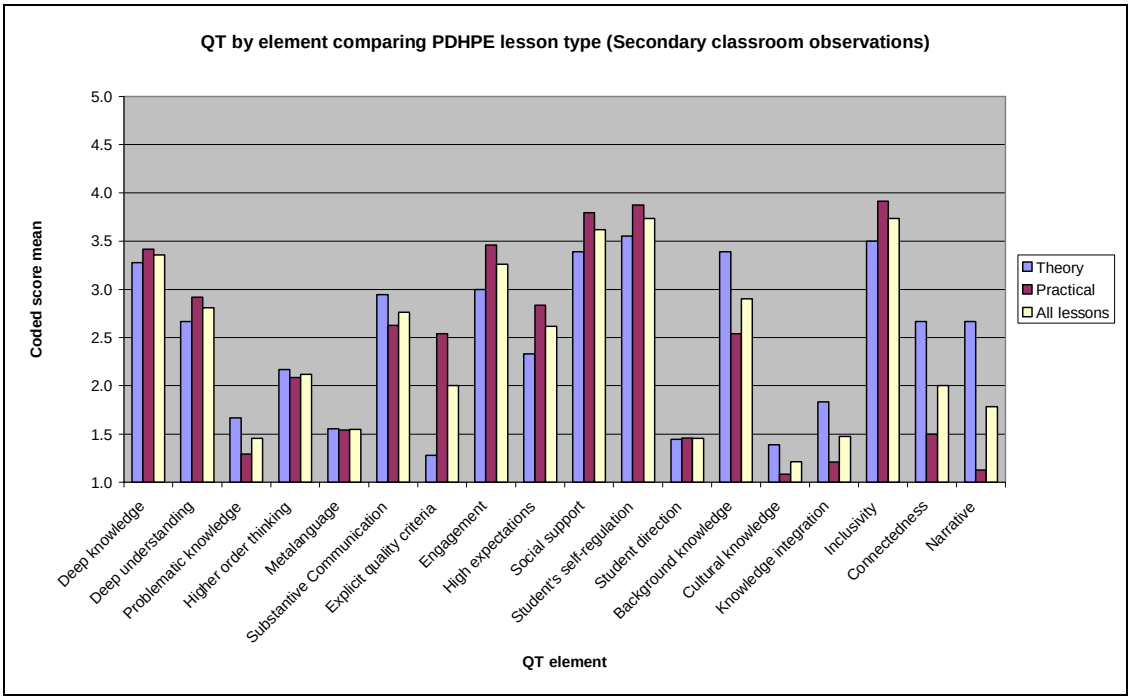


Figure 3. Theory v Practical lessons in PDHPE by QT element

In terms of the Quality Learning Environment dimension, while scores for most elements were slightly higher in practical lessons than in theory lessons, with the exception of Student Direction, the only statistically significant difference was for the element Explicit Quality Criteria. Hence, although practical lessons are sometimes seen by non-PE teachers to be particularly demanding in terms of managing student behaviour, and while the playing field is commonly seen to be a site of bullying or exclusionary behaviour among students, we found no significant differences – indeed, slightly higher mean scores were obtained for Student Self-Regulation, Engagement, Social Support, and High Expectations in practical lessons. This may in part be related to institutional and regulatory restrictions on the subject which mean it is allocated limited time during the students' week. As one of the teachers in our study put it, "with the lack of time that we have available, we expect self-discipline from the students and we do remind them regularly that we only see you once a week for practical and once a week for theory, so let's push on" (794016). The major, statistically significant, difference between practical and theory lessons in the Quality Learning Environment dimension was for Explicit Quality Criteria (*df* 1.0, $F = 16.86$, $p < 0.00$, $\omega^2 = 0.26$), with much higher scores found in practical lessons[2]. These omega squared scores (ω^2) indicate that, for instance, in this case, 26% of the difference in Explicit Quality Criteria scores can be accounted for by the difference between theory and practical lessons.

Even greater differences between practical and theory lessons were found in the Significance dimension. Background Knowledge (*df* 1.0, $F = 7.32$, $p < 0.01$, $\omega^2 = 0.12$), Knowledge Integration (*df* 1.0, $F = 16.96$, $p < 0.00$, $\omega^2 = 0.26$), Connectedness (*df* 1.0, $F = 20.65$, $p < 0.00$, $\omega^2 = 0.30$), Narrative (*df* 1.0, $F = 35.18$, $p < 0.00$, $\omega^2 = 0.43$), and Significance overall (*df* 1.0, F =

[2] Indeed the EQC mean score for practical PDHPE lessons was also higher than the mean for any other subject in our study, as were the means for every other element of this dimension except Student Direction.

23.93, $p < 0.00$, $\omega^2 = 0.34$) were all higher in theory lessons than in practical lessons[3]. Tests of effect size are reported in Table 6, with the higher scoring lesson type depicted in bold. In all cases except Quality Learning Environment overall, these effect sizes measures indicate large differences/associations.

Table 6. Theory v Practical lessons in PDHPE

Elements	Lesson type	n	mean	sd	se	Sig (2-tailed)	Cohen's d	F	ω^2
Explicit quality criteria	theory	18	1.28	0.46	0.11				
	practical	24	2.54	1.10	0.23	0.00	1.50	20.82	0.32
Quality Learning Environment	theory	18	15.00	4.14	0.98				
	practical	24	17.96	3.36	0.68	0.02	0.78	6.54	0.12
Background knowledge	**theory**	18	3.39	1.29	0.30				
	practical	24	2.54	0.66	0.13	0.01	0.83	7.72	0.14
Knowledge integration	**theory**	18	1.83	0.62	0.15				
	practical	24	1.21	0.51	0.10	0.00	1.10	12.91	0.22
Connectedness	**theory**	18	2.67	1.19	0.28				
	practical	24	1.50	0.59	0.12	0.00	1.24	17.50	0.28
Narrative	**theory**	18	2.67	1.14	0.27				
	practical	24	1.13	0.45	0.09	0.00	1.78	36.73	0.46
Significance	**theory**	18	15.44	3.55	0.84				
	practical	24	11.38	2.02	0.41	0.00	1.41	22.12	0.33

These differences in Quality Teaching scores between practical and theory lessons are summarisedin Table 7.

Table 7. Summary of differences between Practical and Theory lessons

	Practical lessons	***Theory lessons***
Intellectual Quality	No statistically significant differences	
Quality Learning Environment	Higher EQC** and overall QLE*	
Significance		Higher BK**, KI**, C**, N** and overall SIG**

* = $p < 0.05$, ** = $p < 0.01$

Given the large number of significant differences in the Significance dimension, possible explanations for the lower Significance scores in practical lessons are explored below:

Background Knowledge might be less frequently drawn upon in practical lessons given the typically wide range of experience among students in each practical activity in terms of levels of previous engagement and performance. That is, classes can have the whole range of backgrounds from complete novices to representative players. In the interests of giving all students a chance to succeed, teachers might be careful to introduce activities and concepts as if encountering students

[3] Note also that Connectedness and Narrative in practical lessons were also lower than the mean scores for any other subject, while Background Knowledge and Knowledge Integration were higher in theory lessons than they were in any other subject.

who are new to the knowledge being presented. The high scores on Explicit Quality Criteria attest to teachers' explicitness in practical lessons as they attempt to help all students understand what is required to perform well. In theory lessons, on the other hand, especially given the immediate relevance of many topics to students' lives, it is not surprising to find teachers making more direct links to students' prior experiences.

The lower scores for practical lessons on *Knowledge Integration* are not surprising, given that theory lessons in PDHPE provide plenty of opportunities to connect with learning in such subjects as HSIE, Science, Maths and/or English. Such opportunities would seem less common, though not inconceivable, as teachers instruct students in the skills of volleyball or strategies in game play, particularly when, as already demonstrated, the approach to such lessons was quite traditionally and narrowly focused on learning game skills. One inexperienced teacher in our study conveyed a strong sense of the limits to integration for PDHPE anyway. In his view "I can't be teaching literacy and numeracy, which they want us to teach the kids. I mean our subject is not maths; our subject is not English; our subject is Physical Education" (794101).

Given that *Connectedness* is a measure of the extent to which justifications for lessons are made explicit in relation to something outside of the classroom, it is also not surprising to find these lower scores in practical lessons. Again, given that the theory lessons we observed addressed such topics as personal safety, the effects of smoking, and the effects of puberty on the female body, connectedness to students' own lives and to society were often clear. As one teacher put it "I think relevancy is a big issue in PDHPE because the issues that we talk about and the kids, what they want to know about, is highly relevant to them and we try to base our work around … things that are relevant to them" (825015). In many cases, connectedness in practical lessons would be much harder to establish, especially if using the traditional skills-based approach to teaching games that we observed. Indeed, the lower scores for the Connectedness of practical lessons attest to the dominance of traditional approaches to teaching practical lessons in PE despite decades of reform and debate (e.g., see Teaching Games for Understanding, Games Sense, etc). Of the 36 interviews conducted with PE teachers, only one teacher mentioned these alternative approaches:

> And there was a – I forget the name of it now – it was a way of building up a games unit from the bottom up and having the kids initiate the rules and build the rules up slowly. There's a name for it and I can't think of it. I can't think of the guy who did it. We went to Sydney for the training and I think there were four university lecturers … yep, Game Sense it was called. (734078)

This rare reference in our interview data does not suggest a deep understanding of this approach to teaching games, but instead highlights the dominance of traditional skills-based approaches to games teaching.

The lower scores in *Narrative* for practical lessons are also not surprising given that stories are frequently used to illustrate points in theory lessons of the type mentioned above. While stories are not absent from practical lessons, it seems that in general there is less talk and more doing. The slightly lower score for the Substantive Communication element provides some indication that this might be the case.

On the status and quality of PDPHE

The specificity of analysis in this study, made possible by the Quality Teaching model and associated research instruments, has enabled a number of insights to be drawn about the nature and quality of teaching in PDHPE, at least in the NSW public schools that were part of this study.

Some of the findings here are worth celebrating among status-conscious physical educators. That is, despite PE's relatively marginal status among school subjects, our measures of the intellectual demands, learning conditions, and significance of PE lessons and tasks indicate that PE teachers can hold their heads high among colleagues from other subject areas. The positive findings include the high quality of PE tasks in comparison with task scores in 'core' curriculum areas. These scores are indicative of the shift from PE as practical activity to PE as both theory and practical. One teacher in our study described this movement in the subject this way: "The emphasis has changed in that now we are more accountable as far as intellectual stimulus goes. What was once considered an active area, out there and play a game, a lot of it, is now the theoretical background towards it" (825038). The relatively high scores of both theory and practical lessons can also be interpreted as a positive sign, and evidence which responds to critics like Ted Sizer (2004) who once asserted that "physical education is neither" (p. 134). Our data indicate that it is both.

On the other hand, despite PDHPE's relative superiority over other subjects in some elements and dimensions of Quality Teaching, we would caution against smugness or any overblown sense of the relative worth of the subject or the teaching within it. In short, the scores we found in PDHPE could be much higher than they were. That is, even for dimensions where PDHPE has scored highest among subjects in our study, in many instances these scores are not very high in terms of the full range of available scores using the Quality Teaching instruments. That is, the best tasks scored 44-54 from a possible range of 14 to 70 and the best lessons scored 53-68 from a possible range of 18 to 90. Given that the instruments are designed to maximise variability, 'perfect' scores are not anticipated. Nonetheless these data indicate that there is room for improvement in the quality of PE lessons and assessment tasks. Our data on the nature of lessons and tasks also give weight to recent arguments about Physical Education's intransigent character (Penney, 1998) despite disciplinary and contextual shifts in curriculum and approaches to knowledge.

While the Quality Teaching model has been referred to in this paper primarily in relation to the research methodology, the model also has significant implications for professional development. Indeed, the potential impact of Quality Teaching for enhancing teaching in PDHPE was captured in one teacher's declaration that Quality Teaching is:

> making a huge difference in terms of what kids actually walk away with. No longer are we just delivering, I think we're now making them a part of what we are doing and I think that's really quite significant. I think we are… and whether it's historically part of what people's perception of our subject was, being "bats and balls, here's a ball, go and play with it", and I think now that we are asking kids to think about what they can do and what they can contribute to these things – their own health, their own physical activity, then I think that's been far more significantly recognised by kids and therefore the profession as well. (825049)

The data presented in this paper are unlikely to have a major impact on the perceived status of PE. Shifting enduring perceptions, dominant in schools and society, will take more than evidence of the relative quality of Physical Education practice. However, we trust that our analyses deflate some of the challenges against Physical Education as a legitimate and worthwhile school subject while simultaneously inflating both the confidence and the aspirations of physical educators to continue refining curriculum and improving pedagogy in Physical Education.

Acknowledgements

The authors would like to thank sincerely all the teachers and principals who assisted in the conduct of this project. Their willingness to participate in this study is greatly appreciated. The study was jointly funded by the Australian Research Council (ARC) and the NSW Department of Education and Training (2004-2007) as part of the ARC Linkage Program. The views expressed in this paper are the views of the authors. We thank Hywel Ellis for his assistance with the preparation of this paper.

References

Amosa, W., Ladwig, J. G., Griffiths, T., & Gore, J. M. (2007). Equity effects of Quality Teaching: Closing the gap [Electronic Version]. *Australian Association for Research in Education Annual Conference*. Retrieved 23 June 2008 from http://www.aare.edu.au/07pap/amo07284.pdf.

Apple, M. (1979). *Ideology and Curriculum.* London: Routledge and Kegan Paul.

Cohen, J. (1977). *Statistical power analysis for behavioral sciences* (revised edition). New York: Academic Press.

Education Queensland. (2001). *The Queensland School Reform Longitudinal Study*. Brisbane: School of Education, The University of Queensland.

Green, K. (2000). Exploring the Everyday 'Philosophies' of Physical Education Teachers from a Sociological Perspective. *Sport, Education and Society, 5* (2), 109-129.

Griffiths, T., Gore, J. M., & Ladwig, J. G. (2006). Teachers' fundamental beliefs, commitment to reform, and the quality of pedagogy [Electronic Version]. *Australian Association for Research in Education Annual Conference.* Retrieved 23 June 2008 from http://www.aare.edu.au/06pap/gri06386.pdf.

Gore, J. M., & Ladwig, J. G. (2006). Professional development for pedagogical impact [Electronic Version]. *Australian Association for Research in Education Annual Conference.* Retrieved 23 June 2008 from http://www.aare.edu.au/06pap/gor06389.pdf.

Gore, J. M., Ladwig, J. G., Griffiths, T., & Amosa, W. (2007). Data-driven guidelines for high quality teacher education [Electronic Version]. *Australian Association for Research in Education Annual Conference.* Retrieved 23 June 2008 from http://www.aare.edu.au/07pap/gor07285.pdf.

Kirk, D. (1992). *Defining Physical Education: The Social Construction of a School Subject in Postwar Britain.* Falmer: London.

Kirk, D. and Tinning, R. (1990). Introduction: Physical Education, Curriculum and Culture, In D. Kirk and R. Tinning, eds. *Physical Education, Curriculum and Culture: Critical Issues in Contemporary Crisis* (pp.). London: Falmer.

Kirk, R.E. (2007). Effect magnitude: A different focus. *Journal of Statistical Planning and Inference,137* (5), 1634-1646.

Ladwig, J.G. (1996). *Academic Distintions: Theory and Methodology in the Sociology of School Knowledge.* London and New York: Routledge.

Ladwig, J.G., Smith, M., Gore, J. M., Amosa, W., & Griffiths, T. (2007). Quality of pedagogy and student achievement: Multi-level replication of authentic pedagogy [Electronic Version]. *Australian Association for Research in Education Annual Conference.* Retrieved 23 June 2008 from http://www.aare.edu.au/07pap/lad07283.pdf.

Macdonald, D. (1995). The Role of Proletarianization in Physical Education Teacher Attrition. *Research Quarterly for Exercise and Sport, 66* (2), 129-41.

Marshall, J. and Hardman, K. (2000). The State and Status of Physical Education in Schools in International Context. *European Physical Education Review,* 6, 203-229

Penney, D. (1998). School Subjects and Structures: reinforcing traditional voices in contemporary 'reforms' of education. *Discourse: Studies in the Cultural Politics of Education, 19* (1), 5-17.

Newmann, F. M., et al. (1996). "Authentic pedagogy and student performance." *American Journal of Education, 104* (4), 280.

Sizer, T. R. (2004) *Horace's compromise--the dilemma of the American high schoo* : the first report from A study of American high schools, co-sponsored by the National Association of Secondary School Principals and the Commission on Educational Issues of the National Association of Independent Schools. Boston: Houghton Mifflin.

Smyth, D.M. (1995). First-Year Physical Education Teachers' Perceptions of Their Workplace. *Journal of Teaching in Physical Education, 14* (2),

Sparkes, A. and Templin, T. (1992). Life Histories and Physical Education Teachers: Exploring the Meanings of Marginality, in A. Sparkes, ed. *Research in Physical Education and Sport: Exploring Alternative Visions* (pp.). London: Falmer.

Appendix 1: The Quality Teaching model, coding scale questions for classroom practice

Dimensions		Elements	Coding scale question
	Intellectual Quality	Deep knowledge	To what extent is the knowledge being addressed focused on a small number of key concepts and the relationships between and among concepts?
		Deep understanding	To what extent do students demonstrate a profound and meaningful understanding of central ideas and the relationships between and among those central ideas?
		Problematic knowledge	To what extent are students encouraged to address multiple perspectives and/or solutions? To what extent are students able to recognise knowledge as constructed and therefore open to question?
		Higher-order thinking	To what extent are students regularly engaged in thinking that requires them to organise, reorganise, apply, analyse, synthesise and evaluate knowledge and information?
		Metalanguage	To what extent do lessons explicitly name and analyse knowledge as a specialist language? To what extent do lessons provide frequent commentary on language use and the various contexts of differing language uses?
		Substantive communication	To what extent are students regularly engaged in sustained conversations (in oral, written or artistic forms) about the ideas and concepts they are encountering?
	Quality learning Environment	Explicit quality criteria	To what extent are students provided with explicit criteria for the quality of work they are to produce? To what extent are those criteria a regular reference point for the development and assessment of student work?
		*Engagement	To what extent are most students, most of the time, seriously engaged in the lesson? To what extent do students display sustained interest and attention?
		High expectations	To what extent are high expectations of all students communicated? To what extent is conceptual risk taking encouraged and rewarded?
		*Social support	To what extent is there strong positive support for learning and mutual respect among teachers and students and others assisting students' learning? To what extent is the classroom free of negative personal comment or put-downs?
		*Students' self-regulation	To what extent do students demonstrate autonomy and initiative so that minimal attention to the disciplining and regulation of student behaviour is required?
		Student direction	To what extent do students exercise some direction over the selection of activities related to their learning and the means and manner by which these activities will be done?
	Significance	Background knowledge	To what extent do lessons regularly and explicitly build from students' background knowledge, in terms of prior school knowledge, as well as other aspects of their personal lives?
		Cultural knowledge	To what extent do lessons regularly incorporate the cultural knowledge of diverse social groupings?
		Knowledge integration	To what extent do lessons regularly demonstrate links between and within subjects and key learning areas?
		*Inclusivity	To what extent do lessons include and publicly value the participation of all students across the social and cultural backgrounds represented in the classroom?
		Connectedness	To what extent do lesson activities rely on the application of school knowledge in real-life contexts or problems? To what extent do lesson activities provide opportunities for students to share their work with audiences beyond the classroom and school?
		Narrative	To what extent do lessons employ narrative to enrich student understanding?

* The elements of Engagement, Student's self regulation, Social support, and Inclusivity are not measurable in the coding of written assessment tasks.

Appendix 2: Excerpt from the instrument ‘A classroom practice guide’.
Coding scales for Deep knowledge and Explicit Quality Criteria.

1.1 Deep knowledge

Description

Knowledge is deep when it concerns the central ideas or concepts of a topic, subject or KLA and when the knowledge is judged to be crucial to the topic, subject or KLA. Deep knowledge is evident when either the teacher or the students provide information, reasoning or arguments that address the centrality or complexity of a key concept or idea, or when relatively complex relations are established to other central concepts.

Knowledge is shallow or superficial when it does not concern significant concepts or key ideas of a topic, subject or KLA, or when concepts or ideas are fragmented and disconnected from a central focus. Knowledge is also shallow when important ideas are treated superficially by the teacher or students, or when there is no clear focus on an important idea or concept. This superficiality can arise from trying to cover large quantities of fragmented information that results in the content covered remaining unconnected to central ideas or concepts.

Coding scale

To what extent is the knowledge being addressed focused on a small number of key concepts and the relationships between and among concepts?

Deep knowledge

1. Almost all of the content knowledge of the lesson is shallow because it does not deal with significant concepts or ideas.
2. Some key concepts and ideas are mentioned or covered by the teacher or students, but only at a superficial level.
3. Knowledge is treated unevenly during instruction. A significant idea may be addressed as part of the lesson, but in general the focus on key concepts and ideas is not sustained throughout the lesson.
4. Most of the content knowledge of the lesson is deep. Sustained focus on central concepts or ideas is occasionally interrupted by superficial or unrelated ideas or concepts.
5. Knowledge is deep because focus is sustained on key ideas or concepts throughout the lesson.

2.1 Explicit quality criteria

Description

High explicit quality criteria is identified by frequent, detailed and specific statements about the quality of work required of students. Explicit quality criteria become reference points when the teacher and/or students use the criteria to develop and check their own work or the work of others.

Low explicit quality criteria is identified by an absence of written or spoken reference to the quality of work expected of students. Reference to technical or procedural requirements only (such as the number of examples, length of an essay or the duration of a presentation) is not evidence of explicit quality criteria.

Coding scale

To what extent are students provided with explicit criteria for the quality of work they are to produce? To what extent are those criteria a regular reference point for the development and assessment of student work?

Explicit quality criteria

1. No explicit statements regarding the quality of work are made. Only technical and procedural criteria are made explicit.
2. Only general statements are made regarding the desired quality of the work.
3. Detailed criteria regarding the quality of work are made explicit during the lesson, but there is no evidence that students are using the criteria to examine the quality of their work.
4. Detailed criteria regarding the quality of work are made explicit or reinforced during the lesson and there is evidence of some students, some of the time, examining the quality of their work in relation to these criteria.
5. Detailed criteria regarding the quality of work are made explicit or reinforced throughout the lesson and there is consistent evidence of students examining the quality of their work in relation to these criteria.

Appendix 3: Excerpt from the Student Work Coding Scales.

ITEM 1. PROBLEMATIC KNOWLEDGE

Description

Presenting *Problematic knowledge* involves an understanding of knowledge not as a fixed body of information, but as being constructed. Beyond interpretations of and varying perspectives in PDHPE, knowledge about health, personal development and physical education itself is subject to political, social, cultural influences and implications. That is, knowledge is based on collectively unconscious (i.e., cultural) assumptions and the value placed on knowledge has been developed through interactions between people (i.e., social) and power relations (i.e., political). *Problematic knowledge* is evident in student performances in PDHPE when students, for example: provide more than one solution to a problem in adolescent health; analyse conventional understandings of health in relation to the political, social, cultural and/or ethical contexts in which they were generated; substantively criticise and challenge the assumptions that have guided current debates in PDHPE; or, create new (for them) health, personal development or physical education programs.

Knowledge as given sees the subject content within the student performance represented as facts or as a body of truth. The transmission of the information may vary, but is based on the concept of knowledge as being static and not open to interpretation. The student performance attempts to represent a "correct" response, rather than substantiating alternative conclusions. *Knowledge as given* is evident in student performances in PDHPE when students, for example, reproduce predictable perspectives on PDHPE or report alternative perspectives without analysing the knowledge on which these findings are constructed.

NOTE: ***For the purposes of scoring this item, the focus is on the content of the student performance, and a judgement as to the proportion of the presented knowledge that is problematic.***

TO WHAT DEGREE IS KNOWLEDGE PRESENTED AS PROBLEMATIC?

①	Student performance treats no knowledge as problematic. All knowledge is presented in an uncritical fashion as not open to interpretation.
②	Student performance treats minimal amounts of knowledge as problematic, by acknowledging changing or alternative perspectives within PDHPE. Ultimately, findings and conclusions are linked or reduced to a body of facts.
③	Student performance treats PDHPE knowledge as open to social influence to some extent. Student performance demonstrates an understanding of the construction of PDHPE knowledge by demonstrating the bases of alternative positions in PDHPE debates.
④	Student performance treats PDHPE knowledge as problematic and open to social, political and historical influence. Student performance criticises PDHPE knowledge, demonstrating the interrelation between processes, technologies, and knowledge claims in PDHPE within historical and ethical contexts.