Quality of pedagogy and student achievement: multi-level replication of authentic pedagogy[1]

By,
James G Ladwig, The University of Newcastle
Max Smith, New South Wales Department of Education and Training
Jennifer Gore, The University of Newcastle
Wendy Amosa, The University of Newcastle, and
Thomas Griffiths, The University of Newcastle

Paper presented at the Australian Association for Research in Education conference,
Fremantle, 25 – 29 November, 2007.

ABSTRACT

This paper presents SIPA's initial, cross-sectional, analysis of the relationship between pedagogy and student outcomes, when measured in terms of the Quality Teaching model and in-class student performance respectively. Following the school effects tradition, the analysis is based on multi-level modelling in which students' prior achievement, socio-economic status, gender, race and NESB status are taken as background variables.

Measures for pedagogy are based on classroom observations and task coding of the three dimensions of pedagogy defined in the Quality Teaching model, Intellectual Quality, Quality Learning Environment and Significance and the Authentic Pedagogy constructs on which the Quality Teaching model was based. The outcome variable in this modeling is Newmann and Associates' (1996) Authentic Achievement scale, applied to student work samples gathered from observed classes.

This is the first quantitative analysis of the efficacy of the Quality Teaching model for predicting student outcomes and has implications for the viability of the model and its predecessors. In essence this is a replication and expansion of the work reported by Newmann, Marks and Gamoran (1996), Newmann, Lopez & Bryk (1998) and Newmann, Bryk, & Nagaoka (2001). As such, it offers the first rigorous test of the cross-national transportability of Authentic Pedagogy.

[1] As part of a symposium, this paper should be read in conjunction with the other papers presented; the complete series is: GRI07282, LAD07283, AMO07284 and GOR07285.

Introduction

More than a decade ago, in 1996, Fred Newmann and his colleagues began reporting the results of studies designed to test the efficacy of a set of pedagogical standards Newmann termed 'Authentic Pedagogy' (Newmann, Marks and Gamoran, 1996; Newmann and Associates, 1996). In the midst of a popular push toward several forms of 'constructivist' teaching methods, Newmann's agenda was in many ways a corrective to the desire to expand the deployment of 'active' forms of teaching. That is, although Newmann's notion of Authentic Pedagogy was consistent with, and generally sympathetic to, 'constructivist' teaching, the research and researchers led by Newmann were acutely aware and generally in agreement that many forms of pedagogy intended to provide students with the opportunity to actively construct their understanding of matters at hand would simply miss the mark. However well intended, Newmann et al identified the problem with the blanket advocacy of 'active' pedagogy as a need to identify standards for the intellectual quality since without them there is a risk that:

> … efforts focused on active learning may lead down an illusory path where student participation in activities can become an end in itself, regardless of the intellectual quality of students' work (Newmann, Marks and Gamoran, 1996: 281).

While there was consternation among some critics about the notion of 'authenticity,' even the staunchest critics of authentic pedagogy have acknowledged the rigour and import of Newmann's research in establishing a strong link between Authentic Pedagogy, as defined by Newmann, and student performance (see, e.g. Terwilliger, 1998). There has been a substantial amount of subsequent research supporting the original analyses of Authentic Pedagogy, but none of that has been replicated in Australia. The main purpose of this paper is to report the findings of the first replication of a test of Authentic Pedagogy in Australia.

Significance of the study

Since its initial reporting, the model of Authentic Pedagogy identified by Newmann and Associates has been examined in three forms of analyses: descriptive, cross-sectional, and longitudinal. These studies into the effects of Authentic Pedagogy have been conducted in relation to mathematics, English, the social studies, and science curricular areas, and across the primary, middle and secondary years of schooling.

First, descriptive studies of the prevalence of 'intellectual quality' have documented the relative scarcity of pedagogy focused on high levels of intellectual quality. These descriptive studies have measured both in-class instructional practices, employing direct observation techniques and indirect survey instruments, and the quality of in-class assessment practices, from samples of assessment tasks employed by teachers, in the US (Lee and Smith, 1993; Lee and Smith, 1995; Newmann, Mark, Gamoran, 1996; Newmann and Associates, 1996; Newmann, Lopez and Bryk, 1998), in Australia (Gore, Ladwig, Lingard, Luke et al, 2001), and in the Netherlands (Reolofs and Terwal, 1999). Thus, using a variety of measurement techniques, these studies have collectively demonstrated that high levels of intellectual quality are not common in most classrooms.

Another descriptive finding in the original Newmann and Associates analyses was that the distribution of pedagogy was equitable, in the sample of restructured schools used for the initial studies (see Newmann and Associates, 1996; Newmann, Marks and Gamoran, 1996). That is, there was very little difference in the extent Authentic Pedagogy was found in schools and classrooms with high and low percentages of

racial and ethnic disadvantaged groups and no bias in terms of social class difference (all classes were co-educational). This was an important factor in analyses of the effects of Authentic Pedagogy for these equity groups in that any attempt to find effects for disadvantaged groups requires having it present in the first place. Finally, still at a descriptive level, as is now commonly found in decomposition analyses of school effects, there were relatively large variances of Authentic pedagogy within schools, between classes, as opposed to much smaller differences at the between school level (see Newmann, Marks, Gamoran, 1996).

Second, employing single point in time data analysed with multilevel, multivariate models, cross-sectional 'effect' studies have been able to document a strong association between Authentic Pedagogy and improved student outcomes after controlling for the independent effects of students' prior achievement, students' social backgrounds and several school level organisational, demographic and pedagogical variables. Strong positive effects have been found in studies that measure student outcomes by applying standard 'authentic' criteria to student performances on in-class assessment tasks (where tasks differ from class to class). This work, led by Newmann, measured pedagogy by coding both classroom observation and in-class assessment practices (Newmann, Marks, Gamoran, 1996; Newmann and Associates, 1996), and by coding in-class assessment practices alone (Newmann, Lopez and Bryk, 1998; Newmann, Bryk, and Nagaoka, 2001) . Avery (1999) extended these studies by examining the effects of pedagogy (measured via classroom observation) on student performances on a common performance-based task, with similar results. Importantly, one of these cross-sectional studies found comparable positive effects of Authentic Pedagogy on student outcomes when measured on a conventional standardised test of basic skills (Newmann, Bryk and Nagaoka, 2001).

Third and finally, using data available from the US National Education Longitudinal Survey (NELS), Lee and Smith (1995) and Lee, Smith and Croninger (1997) have been able to estimate the effects of 'authentic' pedagogy on student gains in national criterion-referenced tests in mathematics and science. Here it is also important to note that the Authentic Pedagogy construct was primarily focused on intellectual quality. Parallel studies of the effects of 'supportive classroom environments' (the US construct most closely related to the NSW 'quality learning environments') have shown independent efficacy of that construct on improved student learning outcomes (see, e.g. Newmann and Associates, 1996).

Taken together, these past studies provide a substantial case for focusing on the possible similar effects of Intellectual Quality as defined in Authentic Pedagogy in Australia. While the Queensland School Reform Longitudinal Study, directed by Ladwig and Lingard, attempted to develop a replication of this work, lack of prior achievement measures of the students in that Queensland sample meant that a full replication and rigorous test were not possible in the QSRLS (See Ladwig, 2007). Never-the-less, the potential offered by the Authentic Pedagogy model has led to substantial professional development work in many states around Australia. The task for this analysis is to replicate some of the first models developed by Newmann and his associates, to test the viability of that professional development work for improving student performances. In short hand, the central question for this analysis is, are the standards of intellectual quality in Authentic Pedagogy associated with improved student performance, above and beyond standard background control measures? None of the studies that analyse the effect of authentic pedagogy on student outcomes were conducted in Australia. Such a cross-national view is vital before even larger investments are made in models based on Authentic Pedagogy,

and before much more ungrounded debate about such models continues in political circles. This analysis is the first direct replication to do so.

Design for this analysis

As noted above, the original analytical constructs employed in the Authentic Pedagogy studies have by now been incorporated into several state level reform initiatives, beginning with Queensland's New Basics initiative but continuing through several other states' reforms. More importantly for this study, however, is that in addition to the analytical constructs, the original research instruments have been adopted in both Queensland and NSW as a basis for their own pedagogical professional development material. The present study employs the instrumentation developed by Ladwig and Gore for NSW-DET, included in which were the same items employed by Newmann and Associates in the original Authentic Pedagogy research. In other words, the same instruments used by Newmann have been replicated in this analysis, with only slight variation in the nomenclature. (NSW-DET wisely altered some of the naming of items to better match its needs.)

The shared analytical construct includes the same three aspects of pedagogy employed in Authentic Pedagogy, those Newmann et al named, 'construction of knowledge', 'disciplined inquiry, and 'value beyond schools'. For the first studies of Authentic Pedagogy, Newmann et al conceived of pedagogy to include both classroom instruction and in-class assessment tasks (e.g. Newmann, Marks and Gamoran, 1996). In that study, each class included in the study was observed a total of four times in two field visits in one year (fall and spring), where lessons were coded using an observational instrument designed to measure Authentic Instruction. The outcome measure in that study was based on student work samples (whole classes) collected twice a year (fall and spring) along with the assessment task on which the student work was based. The student work was coded using an instrument designed to measure 'Authentic Achievement', and the task coded with an instrument designed to measure 'authentic assessment'. For the Newmann and Associates (1996) and the Newmann, Marks and Gamoran (1996) analyses, the classroom observation scores and the task scores were combined into an overall additive scale for the Authentic Pedagogy measure.

In this way, the tasks were directly related to the curriculum delivered by teachers (since they designed and were using the tasks), thus sustaining 'curricular validity'. The observation scores were less directly related to the student work, as they did not necessarily relate to the specific unit from which the tasks were taken. In later studies, in Chicago, Newmann relied solely on tasks scores to represent Authentic Pedagogy (see, e.g. Newmann, Bryk and Nagaoka, 2001). This decision was motivated by several factors, including budgetary constraints and the relative contribution of observation scores in regression analyses, when combined with teacher designed tasks.[2]

In addition to following the logic of Newmann's original analyses, several other factors impacted on the modelling choices for the analyses reported here. First, there was a notable tendency among teachers within schools to employ shared task. This reduced the within school unit variation for each time period, each data collection point. Additionally, in the study from which this data is drawn, it was not uncommon for there to be low compliance in each collection point in several schools (not in all collections, nor necessarily consistently by school). Consequently, for any

[2] Personal communication with Newmann and Newmman, Marks, Gamoran, 1996: 309, fn 19.

one collection point there are several schools in the sample for which there was an insufficient number of within school units (tasks) to carry out a three-level model (student within task within school). Thus, for this analysis, we employed a two-level model of students within tasks.

While there are many limitations implied in this design, the approach was based on a need to establish 'existence proof'. That is, the original cross-sectional design was intended to establish the simple issue of whether or not pedagogy that exhibited the qualities of the Authentic Pedagogy model was possible and if possible whether or not it was associated with higher quality student performances, 'above and beyond' controlled background variables. These same questions hold for the Australian context, thus the design of the current analysis mirrors these original Authentic Pedagogy studies as closely as possible.

The data

The data analysed here is drawn from the Systemic Implications of Pedagogy and Achievement in NSW Public Schools study (SIPA). SIPA's overall design is a longitudinal study of the relationships between teachers' professional learning, the quality of pedagogy and student achievement, for three overlapping cohorts of students, over fours years. Schools for SIPA were selected to form a sample, stratified to include primary and secondary, urban and rural school, and schools with high, middle and low SES student populations. All students in the year level of interest were included in the sample within each school. For the current analyses, one collection point of SIPA was selected to mirror the cross-sectional data employed in the replicated studies. This data is from SIPA's second collection of student work in 2005, when students in that sample were in years 4, 6, and 8.

As part of the SIPA design, each participating school submits student work samples (full year level populations) at six time points (one in late 2004, two in 2005 and 2006, one in the beginning of 2007), in two KLAs (out of five possible KLAs with each school completing at least one of English or Maths). Each set of student work is accompanied by the task which generated the students' work. The student work was coded, by teachers in the study, as an 'Authentic Achievement' measure using the Authentic Student Work scales, adopted from Newmann's earlier work (see Newmann and Associates, 1996). This was a three item additive measure, with each item coded on a 1 to 4 Likert scale. Approximately 25% of the student work was double coded with inter-rater reliability checks conducted throughout the coding sessions. Levels of inter-rater reliability were consistent with those reported in previous research.[3] Internal reliability of the scale was also similar to previous research (Cronbach's alpha = .76). For the double coded work, final scores were the agreed common scored that resulted from the inter-rater reliability check. Variance decomposition for the two-level model found that 48% of the variance in student performance was at the task level, 52% at the individual performance level.

The tasks submitted by teachers were also coded using the Authentic Task measure noted above. The model of pedagogy used for SIPA is the NSW Quality Teaching model, but since the items of the Authentic Pedagogy construct are part of the NSW-QT model a direct replication of the Authentic Task measure was possible. This is a seven item additive measure, with each item coded on a 1 to 5 Likert scale in the

[3] For inter-rater reliability, Pearson correlation coefficients and percentage agreements were reported in prior research. For our data, the Pearson correlation coefficients ranged from .52 - .84., exact agreement ranged from 64%- 76%, and within one agreement from 92%-99%.

NSW version. All tasks were at least double coded, with a minimum of one teacher-research pair conducting inter-rater reliability checks on each task, with final scores being the agreed score. Consensus was always possible. Item reliability measures on the tasks scores were also similar to those reported in previous research,[4] with a strong internal consistency measure (Cronbach's alpha = .86).

In addition to these measures of pedagogy (coded tasks in this case) and achievement (coded student work), additional data was drawn from the NSW-DET data bases for the purposes of developing control measures on students' prior achievement and social backgrounds. From schools' data bases it was possible to draw prior achievement results in literacy and numeracy, students' sex, ATSI status, NESB status and SEIFA estimates of social economic status. Data matching was possible for the Year 4 and Year 8 cohort. Prior achievement measures were from these students' Year 3 and 7 Basic Skills results. Matching the studies data with the students' background data yielded a sample from 26 SIPA schools (15 Primary, 11 secondary) who submitted a total of 78 Tasks and student work from 1374 students (640 primary, 734 secondary) in a total of 2236 performances (1119 Year 4 and 1117 in Year 8).

Variables

A summary of the SIPA variables employed in the replication is presented below:

Authentic Achievement Scale: replication of Newmann et al measure with nomenclature changes (3 – 12 range), standardised;
Female: dummy variable, 0 = male, 1 = female;
ATSI (Aboriginal and Torres Strait Islander): dummy variable, 0 = no, 1 = yes;
ESB (English Speaking Background): dummy variable 0 = no, 1 = yes;
SES: scale developed by the ABS for 'census collection districts', areas of 200 households, a couple blocks squared, identified by GPS coordinates, attributed to individual level, standardised (mean of 1000 for all of Australia);
Prior Achievement: mean of prior literacy and numeracy achievement measures developed by DET (logarithmic growth scales for comparable year 3 and year 5 scores), standardised;
Task Average Prior: Prior Achievement average for task group;
Authentic Task Score: Additive scale inclusive of: Deep knowledge, deep understanding, problematic knowledge, analysis, elaborative communication, background knowledge, connectedness (7 – 42 theoretical range), standardised.

Details of the descriptive statistics for the comparable variables employed by Newmann et al can be found in Newmann, Marks and Gamoran (1996). Table 1, below, presented the descriptive statistics for the variables discussed above.

[4] Task Item reliability:

Table 1: Variable Descriptive Statistics (n=2236)

	Minimum	Maximum	Mean	Std. Deviation
Authentic Achievement Scale	3	12	7.24	2.16
Individual SES scale	459.6	1171.36	946.68	98.55
Aboriginal or Torres Strait Islander	0	1	0.09	0.28
English Speaking Background	0	1	0.63	0.48
Sex	0	1	0.50	0.50
Mean Individual Prior Achievement	21.75	80.48	55.72	8.66
Authentic Task Score	7	29	19.58	4.79

Findings

The Quality and Variability of Authentic Pedagogy and Achievement in the SIPA sample

All schools on the SIPA sample had self-nominated for the study at least in part due to an expressed interest in developing or improving the quality of the pedagogy in school. The schools differed in the amount of time prior to the study that each had an explicit commitment to work with the NSW Quality Teaching model in this endeavour. Some schools had been working with the Quality Teaching model (and /or its predecessors) for several years, others had only begun at the beginning of the study in 2004. Each school had explicitly selected two KLAs for inclusion in the study. Thus, the general findings on the level of authentic pedagogy represented in the Authentic Task Score for the data analysed here needs to be considered in this overall trajectory of school-based pedagogical improvement.

The degree of authentic tasks observed is reported in Table 2, below. With and overall range from 7 to 35 on the Authentic Task Scale, the overall mean of 17.62 is lower than the midpoint of 21. Variation between year levels and specific KLAs suggests that significant differences are found in the collected tasks, most notably in primary Maths and in the overall primary/secondary difference. Overall, secondary tasks scored higher than their primary counter-parts on the Authentic Task Scale. (It should be noted that this is not true on classroom observation scores.) Secondary English, PDHPE and Science all scored above the midpoint on the scale as did the single HSIE task submitted for this collection. KLA differences accounted for a substantial portion of Authentic task score variance relative to individual tasks, with a 74% v. 26% differentiation (KLA v individual task within KLA). Similar variance portioning results are obtained when comparing the variance of individual tasks within cohort (Year level), with 70% of that variance being between cohort and 30% between individual tasks within Cohort.[5] On the surface of these variance estimates, both year level and KLA carry substantial influence in determining the overall nature and quality of assessment tasks submitted for the second 2005 SIPA collection.

[5] Variance partitioning estimates of authentic task sores indicates that 18% of the variance in tasks score lies between tasks with 82% between schools.

Table 2: Levels of Authentic Task Score and Student Authentic Achievment by Year Level and KLA

			Authentic Task	Authentic Achievement
Year 4	English	*Mean*	17.95	7.36
		Standard Deviation	3.70	1.86
		Count	22	470
	Maths	*Mean*	14.77	7.00
		Standard Deviation	4.31	2.05
		Count	19	402
	HSIE	*Mean*	18.54	6.98
		Standard Deviation	3.99	2.30
		Count	12	247
Year 8	English	*Mean*	22.90	8.19
		Standard Deviation	3.56	2.14
		Count	15	505
	Maths	*Mean*	17.45	6.27
		Standard Deviation	2.81	1.93
		Count	3	119
	HSIE	*Mean*	27.00	5.93
		Standard Deviation	0.00	1.69
		Count	1	82
	PDHPE	*Mean*	21.70	7.44
		Standard Deviation	2.96	2.01
		Count	4	295
	Science	*Mean*	22.18	5.37
		Standard Deviation	0.39	2.12
		Count	2	116
Total		*Mean*	17.62	7.24
		Standard Deviation	5.15	2.16
		Count	78	2236

The overall levels obtained indicate that high levels of authentic task scores were uncommon and the promotion of intellectual quality as defined by the Authentic Pedagogy construct is clearly a challenge for teachers, even within schools with an overt commitment to improving pedagogy using models that support that construct.[6] These task quality results are quite consistent with prior research on the question of pedagogical improvement generally, and more specifically on the promotion of Authentic Pedagogy (see, eg. Newmann, Mark, Gamoran, 1996: 296).

The quality of student performance is also consistent with prior research (see, Newmann and Associates, 1996). The differences between year levels on student achievement measures is interesting in that for at least two KLAs the average scores

[6] It should be noted that because the sampling tracked students, just because a school identified pedagogical improvement as a goal did not mean that all teachers within the school shared this view.

are actually lower in year 8 than they were in yeah 4 (Math and HSIE). While interesting, it is important to keep in mind that these scores are for different cohorts and would be related to several unaccounted background factors. Attempting to account for these background variables is, of course, precisely the point of the modelling their effects.

On the distribution of pedagogy

One of the more important, but less noted, findings from the original CORS studies was that the extent of Authentic Pedagogy was found to be quite socially equitable across schools, students' social background and levels of prior achievement (Newmann, Mark, Gamoran, 1996). This finding was perhaps not entirely surprising, since many of the restructured schools in the CORS studies had openly, and publicly, been tackling the problem of providing high quality education to traditionally under-achievement populations of students. That these schools had managed to deliver that quality in terms of pedagogy was very good news, but that news essentially indicated that high quality was possible in schools which specifically attempted to provide it. The main surprise was this extended to the classroom, in pedagogy.

In the NSW sample, the same level of equitable distribution of pedagogy does not hold. The SIPA study has well documented significant inequity in the social distribution of pedagogy along several dimensions of student social background, mostly notably ATSI status (e.g., Amosa and Cooper, 2006). The question in all such studies, however, is the degree to which any socially inequitable distribution of pedagogy is related to students' prior achievement. Without commenting on whether or not it is justifiable, at this point, it is widely acknowledged that many teachers would justify socially differential pedagogical expectations in terms of individual prior achievement. However, since the known effects of Authentic Pedagogy in US research were significant and similar for students with both high and low prior achievement (see Newmann, Bryk, and Nagaoka, 2001), the distribution of Authentic Pedagogy in NSW is of substantial interest to our current study.

To examine the distribution of Authentic Pedagogy according to students' prior achievement, aggregate prior achievement scores were generated from individual prior achievement measures at the task (class) level and correlated with our Authentic Pedagogy scores per task. This analysis would reflect, then, the degree to which average class prior achievement relates to the degree to which teacher tasks require student to address the standards of Authentic Pedagogy (deep knowledge, deep understanding, problematic knowledge, analysis, elaborative communication, background knowledge, connectedness). Unlike the CORS sample, however, the SIPA sample revealed a significantly unequal distribution of Authentic Tasks in terms of students' prior achievement (at the task level). The correlation between these two measures ($r = .60$, $N = 78$) was highly significant and positive, indicating that students in classrooms with higher mean prior achievement were given tasks that required significantly higher levels of Authentic Pedagogy. This pattern would be well known to most educators, but was precisely that which the restructured schools in the CORS sample were struggling not to replicate (successfully).

Perhaps most importantly for the current analysis, however, is that the full potential of Authentic Pedagogy can only really be estimated if counterfactual cases exist in the population under study. That is, given the conventional unequal distribution of pedagogy found in the SIPA sample, and given that the correlation between students' prior achievement and task quality was such that cases where the conventional pattern were rare, any estimation of the effect of Authentic Pedagogy in the SIPA sample would be mitigated by the initial distribution of pedagogy found in the first

place. Figure 1, below, graphically demonstrates the problem uncovered in this analysis.

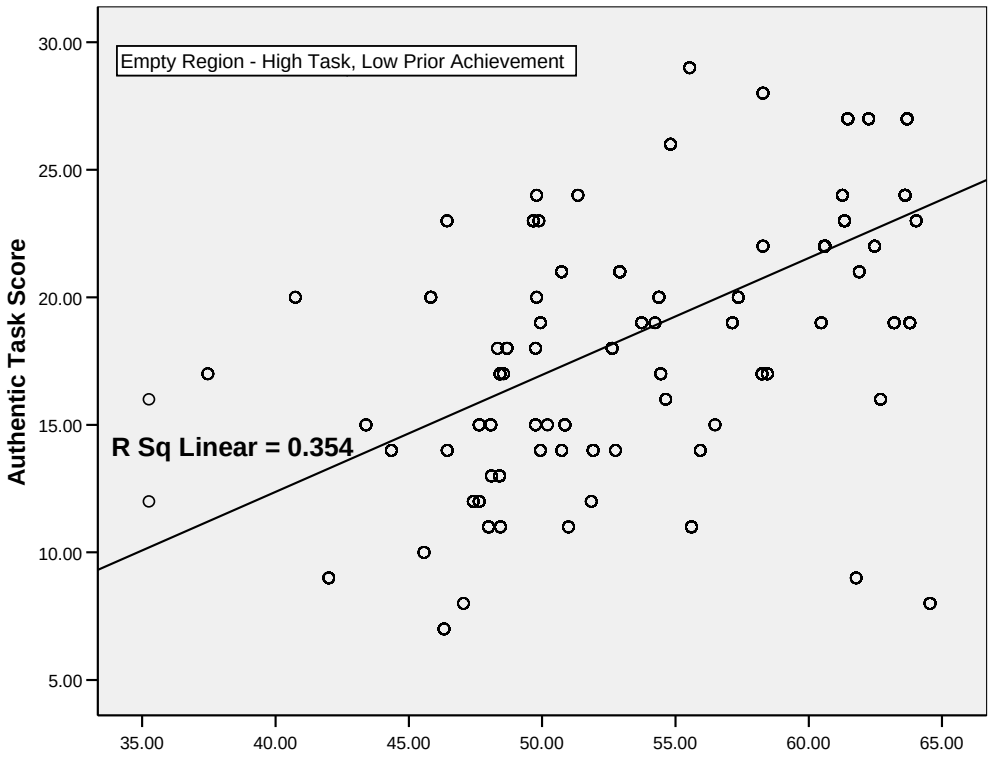

Figure 1: Scatter plot of Task mean prior achievement (x-axis) by Authentic Task Score (y-axis)

In the most ideal sample for this study, it would be possible to analyse their effectiveness for populations of students that would typically not be expected to tackle the challenging standards of Authentic Pedagogy. To estimate this particular effect, however, it is necessary to find at least some example of cases where students of low prior achievement were given tasks that reflect reasonably high scores on the Authentic Task scale. As indicated in the notated empty region of the scatterplot in Figure 1, there were no such cases in the current SIPA sample.

This conventional distribution of pedagogical quality (widely documented in literature on streaming and tracking), carries significant consequences when prior achievement is related to social inequities – which is often the case. Of more immediate concern for the current analyses, however, is that this co-linearity must be kept in mind when interpreting any effects models.

On the effect of Authentic Pedagogy in the SIPA sample

In order to draw direct comparison with the original CORS modelling, the specific fixed effects model developed by Newmann, Marks and Gamoran (1996) was replicated as closely as possible with the SIPA data. The main difference between the models, as noted above, is that the CORS model was a three-level model (student, class, school: with no level 3 predictors) and the SIPA model was a two-level model (student within task). The same predictor variables employed in the CORS model were included in the SIPA modelling. The program used was also different: CORS employs HLM, for SIPA we are using MLWin (v2.02).

Table 3, below, presents the results of the replicated multi-level model along side the original CORS estimates reported in Newmann, Marks and Gamoran (1996). Included in this model are four individual level social background variables (gender, race, language background and socio-economic-status), prior achievement at both the individual and class / task level, the Authentic Pedagogy measures (tasks alone for SIPA) and a cross-product isolating the interaction of prior achievement and task quality.

Table 3 : Multi-level Model estimates (fixed effects)

CORS Model (HLM)		SIPA Replication (MLWin)	
Intercept	-0.07	Intercept	-.08
Female	.10**	Female	.20**
African American	-.16*	ATSI	-.12*
Hispanic	-.10	ESB	-.03
SES	.05	SES	-.03
NAEP achievement	.27**	Prior Achieve	.36**
Class Average NAEP	-.01	Task Average Prior	-.29*
Classroom Authentic Pedagogy	.37**	Authentic Task	.21**
NAEP by Authentic Pedagogy	.08**	Prior by Authentic Task	.05*

** $p < .001$
* $p < .05$

From this table, four differences are notable. First, the estimate of the effect of gender (in favour of girls) is substantially larger in the SIPA sample (about twice as large), where the other social background variables are of similar magnitude. Second, the effect of prior achievement is about a third larger for the SIPA sample and equally significant. Third, the effect of the quality of pedagogy, SIPA's Authentic Tasks measure, is both significant and substantial but lower than the original CORS estimates. For SIPA, one-standard deviation increase in Task quality equates to roughly one-fifth of a standard deviation increase in student outcome, whereas that equivalent increase for CORS was a bit less than two-fifths of a standard deviation increase in authentic achievement. Fourth, the effect of the task average prior achievement in SIPA was significant (whereas it wasn't in CORS), substantial (-.29) and interestingly negative. Each of these findings are worthy of discussion.

Discussion

The relative increase in girls' performance as compared with boys is not new, although the size of the effect in the SIPA sample may warrant further investigation. For the current analyses, we explored whether or not the effect of our pedagogy measure differed for boys as compared to girls (varying the slope); but it did not. We also explored whether adding the interaction between gender and prior achievement altered the effect of either individual factor, with no significant shift in each estimate. Consequently, we conclude for the current study that the gender difference noted above is independent of the quality of pedagogy delivered to girls, as compared to that delivered to boys; and independent of individual student's prior achievement. Beyond this, we simply note that the size of the gender difference would be related to the nature of the outcome measure and the difference between the SIPA and CORS sample may relate to larger cultural differences associated with gender and schooling.

The relatively larger effect of prior achievement in the SIPA sample, as compared with the CORS sample is both notable and concerning. Here the background

distribution of the quality of pedagogy must be kept in mind. This result is perhaps not entirely surprising since it is literally the case that, in this sample, no low prior achievement task group (class) was given a task that required high levels of the Authentic Pedagogy standards. As compared to the CORS sample, where the distribution of pedagogy was more equitable, it is likely that the effect of prior achievement reflected in these findings is part of a much deeper set of relationships that require substantial study.

Here it is important to note that the average prior achievement scores did differ widely. This suggests that there are big differences between groups of students, based on prior achievement, in the SIPA sample. Since the current model does not differentiate between schools and classes, it is not possible to say whether this is more or less related to within school grouping of students. When set aside international studies of streaming and tracking, however, this evidence does provide a strong reason to call for further study of the distribution of the quality of pedagogy between classes. There are currently very few recent studies of streaming and tracking in Australia. We believe these findings make it very clear that there is a strong need to pursue this line of research more vigorously in Australia, as it seems the unequal distribution of the quality of pedagogy may be contributing to a much larger effect of student's individual prior achievement, on subsequent performance, as compared to a replicated US study. Keeping in mind that the US study in question was of the most equitable schools in the US (not the typical US schools), the standard implied here is one that would reflect the current interest in international comparisons that seeks to increase Australia 'equity effects'.

Third, it is important and confirming that the effect of Authentic Task scores, above and beyond students social background and individual prior achievement was both significant and positive. While not as large an effect as that found in the initial CORS study, these findings do suggest that further exploration of the effects of models of pedagogy based on the Authentic Pedagogy model, as are the NSW Quality Teaching and the Queensland Productive Pedagogy models, are warranted. Reasons for the relatively lower effect size, in SIPA as compared to the first CORS analyses, could include: 1) including only task scores in absence of classroom observational scores of pedagogy, 2) the relatively small range of Authentic Tasks scores, 3) the uneven distribution of the quality of the tasks, and 4) employing a two-level model, which potentially underestimates level two effects. Never-the-less, the current findings are encouraging and noteworthy.

Fourth, the significant, substantial and negative effect of average prior achievement requires substantial further analyses. On the surface of it, this finding suggests that the higher the average prior achievement in a class, the lower the achievement – relative to that we would predict based on all other predictor variables (including individual prior achievement). This is a classic 'contextual' effect; but, the direction of the effect is counter-intuitive to what many would expect. It is possible that this result is due to a lack of a school level in the model (wherein between school differences in prior achievement are confounded with within school differences). This is possibly quite an important consideration since the number of small primary schools in the SIPA sample is reflective of the NSW school population, in which small primary schools are not uncommon outside of metropolitan areas. Future studies will be able to incorporate more within school units (across year levels as the study progresses) and more substantially analyse this finding.

Conclusion

Returning to findings of interest from prior Authentic Pedagogy research, where Newmann and his colleagues found large variance between classes, overall levels of Authentic Pedagogy 'fell well below the highest levels of the proposed standards'. Our findings concur – perhaps more so. That is, while the overall range of pedagogical quality found in the SIPA study was not dissimilar to that found in the original CORS analyses, the means and range of Authentic Task scores in this sample of the SIPA study do not provide evidence that high levels of quality pedagogy are common. To the contrary, as in the US, we would suggest that the findings presented above are an indication that the overall quality of pedagogy found in SIPA leave a lot of room for improvement, when the standards of Authentic Pedagogy are the yardstick.

Further, where Authentic Pedagogy did 'pay off', equitably, with robust cross sectional results for all analyses social groups in the CORS analyses, our initial findings seem to indicate similar, if muted, results. Above and beyond students' individual prior achievement and social backgrounds, our findings suggest that higher levels of Authentic Tasks result in higher student performances of a sufficient magnitude to warrant advocacy of the standards of Authentic Pedagogy. These findings represent the first international, cross cultural, confirmation of the efficacy of Authentic Pedagogy. As such, we believe, these findings are of international interests. While their interest for Australia and US should be clear, we would argue that the fact that Authentic Pedagogy did have positive effect in two countries, even after controlling for distributional factors and individual backgrounds, should give reason to query the efficacy of Authentic Pedagogy on a broader scale.

Finally, where the 'reasonably' equitable distribution of the quality of pedagogy found in CORS suggests that it is possible to extend high quality pedagogical expectations across the student population, this level of equitable distribution is not evident in the SIPA sample. Taken together with the effects of students' individual and group level prior achievement, these findings suggest that the distribution of pedagogical quality is a significant area of study in need of great attention in Australia.

REFERENCES

Amosa, W. & S. Cooper (2006). *Schools matter, teachers matter: Exploring the links between teachers' work, school context, and quality teaching*. A paper presented at the Annual meeting of the American Educational Research Association, San Francisco, April 7-11, 2006

Avery, P. G. (1999). Authentic instruction and assessment. Social Education, 65(6), 368-373.

Crowther, F., Kaagan, S. S., Ferguson, M., & Hann, L. (2002). Developing teacher leaders: How teacher leadership enhances school success. Thousand Oaks, CA: Corwin Press.

Fullan, M. (1999). Complexity and the Change Process. In Change Forces: The Sequel, edited by M. Fullan. London: Falmer Press.

Gore, J. M. (2001). Beyond Our Differences: A Re-Assembling of What Matters in Teacher Education. Journal of Teacher Education, 52 (2), 124-135.

Gore, J.M., Ladwig, J.G., Lingard, R., Luke, A. et al. (2001). Final Report of the Queensland School Reform Longitudinal Study. Brisbane: Education Queensland. Full report available from Education Queensland. Synopsis available from: http://www.education.gov.au; authors listed in alphabetical ordering of CIs.

Harris, A, & Hopkins D. (1999). Review Article: Teaching and Learning and the Challenge of Educational Reform. School Effectiveness and School Improvement 10 (2):257-267.

King, M. B. (2002). Professional development to promote schoolwide inquiry. Teaching and Teacher Education, 18, 243-257.

King, M. B., & Newmann, F.M. (2000). Will teacher learning advance school goals? Phi Delta Kappan, 81 (8), 576-580.

King, M.B., Youngs, P., and Ladwig, J.G. (2003). Teacher Learning, Professional Community, and School Organizational Capacity. Australian College of Education Yearbook. Canberra.

Ladwig, J.G. (1998). Authentic school reform. Discourse.

Ladwig, J. G. (2007). Modelling Pedagogy in Australian School Reform. Pedagogies, 2 (2): 1-18.

Lee. V.E. with Smith, J.B. (2001). Restructuring high school for equity and excellence. New York: Teachers College Press.

Lee, V.E., and Bryk, A.S. (1989). A multilevel model of the social distribution of high school achievement. Sociology of Education, 62, 172-192.

Lee, V.E. and Smith, J.B. (1993). Effects of school restructuring on achievement and engagement of middle-grade students. Sociology of Education, 66 (3), 164-187.

Lee, V.E. and Smith, J.B. (1995). Effects of high school restructuring and size on early gains in achievement and engagement. Sociology of Education, 68 (3), 241-270.

Lee, V.E. and Smith, J.B. (1996). Collective responsibility for learning and its effects on gains in achievement for early secondary school students. American Journal of Education, 104(2), 103-147.

Lee, V.E. and Smith, J.B. (1997). High school size: Which works best and for whom? Educational Evaluation and Policy Analysis, 19(3), 205-227.

Lee, V.E. and Smith, J.B. (1999). Social support and achievement for young adolescents in Chicago: The role of academic press. American Educational Research Journal, 36(4), 907-945.

Lee, V.E., Smith, J.B., and Croninger, R.G. (1997). How high school organization influences the equitable distribution of of learning in mathematics and science. Sociology of Education, 70, 128-150.

Louis, K. S. , Kruse, S. D. & Marks, H. M. (1996) Schoolwide professional community in F. M. Newmann & Assoc. Authentic Achievement: Restructuring schools for intellectual quality, San Francisco, Jossey-Bass, 179-203.

Louis, K.S., Marks, H.M. and Kruse, S.D. (1996). Teachers' professional community in restructuring schools. American Educational Research Journal, 33(4), 757-798.

Louis, K. S., & Marks, H. M. (1998). Does professional community affect the classroom? Teachers' work and student experiences in restructuring schools. American Journal of Education, (106) 4, 532-575.

Newmann, F.M. (1993). Beyond common sense in educational restructuring: The issues of content and linkage. Educational Researcher, 51(4), 546-564.

Newmann, F.M. and Associates. (1996). Authentic Achievement: Restructuring school for intellectual quality. San Francisco: Jossey-Bass.

Newmann, F. M., King, M. B., & Rigdon, M. (1997). Accountability and school performance: Implications from restructuring schools. Harvard Educational Review, 67 (1), 41-74.

Newmann, F. M., King, M. B., & Youngs, P. (2000). Professional development that addresses school capacity: Lessons from urban elementary schools. American Journal of Education, 108 (4), 259-299.

Newmann, F.M., Marks, H.M., and Gamoran, A. (1996). Authentic pedagogy and student performance. American Journal of Education, 104(4), 280-312.

Newmann, F. M., Lopez, G., & Bryk, A. S. (1998). The quality of intellectual work in Chicago schools: A baseline report, Chicago: Consortium on Chicago School Research. (Available from: http://www.consortium-chicago.org/)

Newmann, F.M., Bryk, A.S., & Nagaoka, J. (2001). Authentic intellectual work and standardized tests: Conflict or coexistence. Chicago: Consortium on Chicago School Research. (Available from: http://www.consortium-chicago.org/)

Roelofs, E. and Terwel, J. (1999). Constructivism and authentic pedagogy: state of the art and recent developments in the Dutch national curriculum in secondary education. Journal of Curriculum Studies, 31(2), 201-227.

Smith, J., Lee, V. E., & Newmann, F.M. (2001). Instruction and achievement in Chicago elementary schools. Chicago: Consortium on Chicago School Research. (Available from: http://www.consortium-chicago.org/)

Smylie, M. (1995). Teacher learning in the workplace: Implications for school reform. In T.R. Guskey & M. Huberman (Eds.), Professional development in education: New paradigms and practices (pp.92-113). New York: Teachers College Press.

Terwilliger, J.S (1998). Rejoinder: Response to Wiggins and Newmann, *Educational Researcher, 27 (6)*, 22-23.