

## **Modelling Pedagogy in Australian School Reform**

James G Ladwig  
The University of Newcastle

A paper presented at the 2004 Meeting of AARE  
Do not quote without permission of the author

By now it would be common educational knowledge in Australia that models of pedagogy have become something of the flavour of the month. It might also be relatively well known that one of the significant contributions to the now ubiquitous interest in pedagogy, at least north of the Murray, what the model of pedagogy developed by the Queensland School reform Longitudinal Study (hereafter QSRLS), known as Productive Pedagogy<sup>1</sup>. Others may well know that the department of school education in Queensland, Education Queensland, adopted Productive Pedagogy into their teacher professional development repertoire soon after the first annual report of the QSRLS, before its inclusion in the well known New Basics Framework. With the recent reporting of the research and evaluation of the New Basics, it is time, perhaps, to more fully document the research rationale and analysis which led to the formulation of Productive Pedagogy.

This analysis seeks to provide an account of two things. First, it will provide a brief account of the theoretical and policy interest in the development of the Productive Pedagogy model. As the QSRLS was a project funded by Education Queensland, it was designed to serve both research and policy interests. It is important to document how these differing agenda interacted, as there are some important lessons to be learned from the experience of the QSRLS that have not yet been discussed openly. Second, this analysis will offer a first cut at the analysis leading to the formulation of Productive Pedagogy. Since several authors have already made public claims about the construct, its strengths and weaknesses, it is probably a good idea to put the evidence on the table. The data and analysis reported here is actually only the smallest beginning of what will eventually come forth from the QSRLS research, but it is an essential first step. Consequently, the second part of this analysis will attempt to offer reading a data based introduction into the work of QSRLS and a mapping of the subsequent analyses to follow.

### **A brief note on the historical context**

The peculiar history of Productive Pedagogy begins with Queensland Borbidge government's *Leading Schools* initiative, which was largely a school-based management initiative, but which had its own intriguing twists. That is, in addition to its attempt to establish school councils and to give school leaders discretionary authorities, two specific features of the Leading Schools initiative, not normally associated with SBM pushes, were 1) a public proclamation that the point of Queensland's SBM was to improve the pedagogical practices of teachers, and 2) a less public call for transparent evaluation research documenting the change sparked by *Leading Schools*.

What was particularly interesting about the Leading Schools initiative was that its call for improved pedagogy was linked directly to the then recently reported *Authentic Pedagogy*, from the Center on the Organization and Restructuring of Schools (CORS) at the University of Wisconsin – Madison. For the QSRLS, the *Leading Schools*' adoption of Authentic Pedagogy was crucial. At the time, 1997, it was already publicly reported that Authentic Pedagogy had been demonstrated to improve student outcomes, measured in terms directly related to the Authentic Pedagogy construct (Newmann, Marks, and Gamoran, 1996; Newmann and Associates, 1996), and in terms of national criterion-referenced tests in maths and science (Lee et al, 1997). More importantly, Authentic Pedagogy was shown to be equitably distributed and to have similar effects across a range of student social background (Newmann, Marks and Gamoran, 1996). This finding was consistent with the parallel analyses of instructional and organizational effects on student learning in maths and science using data from the National Educational Longitudinal Survey (Lee, Smith and Croninger, 1997). The message behind these studies, that Authentic Pedagogy offers a means to gain both increased school effectiveness *and* equity, was directly incorporated into the rationale and rhetoric of *Leading Schools*.<sup>2</sup>

Here, the opportunity to extend the work of Newmann and his colleagues was directly welded into Education Queensland's interest in evaluative research on the Leading Schools initiative. While the overall structure of the QSRLS was broad, covering a range of research questions on multiple levels of school reform, including a direct analysis of SBM, the interest in classroom practice was central to the overall study.<sup>3</sup> It was this interest that guided the research team to focus first on extending the findings of Newmann and his associates in relation to pedagogy.

### **From Authentic Pedagogy to Productive Pedagogy: a technical and methodological background to Productive Pedagogy**

To understand how and why the Authentic Pedagogy construct was seen to be in need of extending, a bit more needs to be said about just what Authentic Pedagogy was. Built from a decades-long interest in making schools in the US more educative and equitable, Authentic Pedagogy was constructed from prior research into, *inter alia*, student engagement, higher order thinking, and curricular reform *and* as a response to the 1980s 'crisis' in US schools (wherein things were seen to be so bad, that a full-scale 'restructuring' was needed). The 1980s had been a complicated and contradictory time in US education. Where the decade had been brought in with the patently ideological-driven statement that the sad state of school education had made the US *A Nation at Risk*, the intervening 10 yrs had seen a series of school reform initiatives that covered just about every educational-political interest. The question for Newmann and his colleagues, in the formulation of the federal-funded CORS, was a simple one, 'Did any of the reforms of the 1980s actually improve students' learning?'

For Newmann, Authentic Pedagogy was a specific model of pedagogy that theoretically would be needed to produce student performances that met what he and his colleagues presented as high standards of intellectual demand. As Newmann, Marks and Gamoran (1996) framed it, 'Construction of knowledge through disciplined inquiry to produce

discourse, products or performance that have value beyond success in school can serve as a standard of intellectual quality for assessing the authenticity of student performance' (287). The point of the Authentic Pedagogy research was to test the model of pedagogy to see if teaching aligned with these standards did in fact produce student work defined by these standards of intellectual quality. When operationalised in research, CORS included both classroom practices (instruction) and teachers' assessment practices into the overall estimation of the pedagogy students' experienced. For the QSRLS, with a government funded project with very little lead up time, the primary focus was initially on the data to be gathered first: observations of classroom practice.

In their mid-1990s reporting, CORS composed Authentic pedagogy of three features, each of which is implied in the definition of authentic student performances noted above: *disciplined inquiry*, *students' construction of knowledge*, and *value beyond schools* (see Newmann, 1993; Newmann, Marks and Gamoran, 1996; Newmann and Associates, 1996). For classroom observations, these three elements were observed through four observational item scales. Disciplined inquiry was observed in the degree of the (1) 'Depth of knowledge and understanding' and degree of (2) 'substantive communication'. Students' construction of knowledge was observed through the degree to which students were engaged in (3) 'higher order thinking'. And value beyond schools was estimated in terms of the degree to which activities in the classroom were (4) 'connected to the world beyond schools'. Each of these four qualities of classroom practice were observed and rated, along with two other items, on five-point Likert scales. When analyses of the effects of Authentic Pedagogy were conducted, these four scales were combined in a summated scale (which itself was then combined with a measure of Authentic tasks) that described the overall degree to which classroom exhibited Authentic Pedagogy. It is safe to say these were generic (in terms of not being subject specific), global measures (in terms of being overall measures of whole lessons) of overall classroom pedagogy.

While there are many technical reasons to find these instruments and the subsequent findings using them interesting (these are very rough measures, afterall), one of the most obvious conceptual features of this formulation of Authentic Pedagogy is that it was taken to be uni-dimensional. That is, the measures were treated in a manner that assumed them to represent only one dimension of pedagogy. Another issue of keen interest to the QSRLS was that the underlying construct had demonstrated the equity effects it had. That is, there has been ample debate about the effects of different forms of pedagogy in relation to different social groups, and if one form of pedagogy would have been anticipated (by some) to not work for student from a range of social backgrounds it would have been one that relied deep intellectuality and students' construction of knowledge. For all the current hype about hegemonic appeal of 'constructivist' pedagogy, researchers focusing on equity have long been skeptical about forms of pedagogy that clearly represent that which benefits students' from quite academic backgrounds. To the extent that Newmann's Authentic Pedagogy superficially resembles such forms of pedagogy, the positive equity results from Authentic Pedagogy were at least a puzzle (Ladwig, 1998)<sup>4</sup>.

A full discussion of the issues raised here is clearly beyond the scope of this analysis, but one relatively simple point needs to be made clear. That is, whatever the view taken of authentic pedagogy, it is clear that the only explanation open to Newmann and his colleagues was a generalized humanistic appeal to the assumed intrinsic value of authentic work. The reason this is the only explanation available to the authentic pedagogy research relates to the nature of the construct and its data. That is, aside from another very general concern for social support, no other measures of classroom interaction were included in the CORS research. This is, of course, entirely consistent with the theoretical framework developed for the authentic pedagogy research. This is, however, not the only plausible explanation of the authentic pedagogy results; but to investigate other possible explanations, it is necessary to include a wider view of classroom practices. Thus, the first task of the QSRLS was to replicate and test the authentic pedagogy findings within an Australian context, and to simultaneously widen the aperture used when observing classrooms<sup>5</sup>.

This widening of the observational snapshot taken of classroom was constructed in a relatively straight forward analytical manner: the QSRLS constructed a larger observational instruments that included many more items. With one exception, the format and scaling of the Authentic Pedagogy instruments were retained, and all of the items from the CORS research were included. The one exception came from the CORS item *depth of knowledge and understanding*. Embedded within the definition of this item were both the nature of knowledge present and the nature of understanding demonstrated by students. For measurement purposes, these were differentiated in the QSRLS instrument. In addition to the four from the authentic pedagogy measure (now five in the QSRLS version), two other items used by CORS in a separate analysis of social support were also part of the QSRLS observational scheme. The two additional CORS items measures were student *engagement* and *social support*. Each of these were part of an analysis of social support for students and its association with improved student performance (see Newmann and Associates, 1996).

From this point the major concern was to measure qualities of teaching theorized or previously shown to be associated with improved learning for students from traditionally disadvantaged social backgrounds. The first port of call for this search was an instruments already developed by Ladwig (1995), based on Berlak and Berlak's integration of Bernstein's classification and framing analysis and Bourdieu and Passeron's theory of pedagogy as symbolic violence (Berlak and Berlak, 1981). Clearly, both Bernstein's framework and Bourdieu and Passeron's theories are central in the sociology of curriculum, and pivotal in understanding pedagogical relations related to educational inequality. However, the need of the QSRLS was for items that could readily be converted into ratable scales. Berlak and Berlak's (1981) formulation of dilemmas of teaching provides precisely such a possibility, where opposing poles of their dilemmas are taken as the end points of a simple five point ordinal scale. The items included relatively directly from this adaptation of Berlak and Berlak (1981) included: *Knowledge Problematic*, *Knowledge Integration*, *Problem-Based Curriculum*, *Student Control*, *Self-Regulation*, and *Cultural Knowledges*. As the definitions for each of these is well

documented in the publications of Education Queensland, here it is more important to explain a few nomenclature differences between these labels and Berlak and Berlak's.

*Student control* was essentially all four of the Berlak 'control dilemmas' combined into one. This particular set of dilemmas were directly based on Bernstein's analysis of framing, which names four aspects of classroom practices that could be controlled by either teacher or students: the choice of activity, the overall time dedicated to a particular activity, the pace of students' work and the criteria for that work. From earlier pilot work in NSW (Ladwig, 1995), it was known that very little variance would be found if each of these were measured separately; consequently the decision to combine them into one was taken to increase the possibility of finding variance.

Substantially altered, but sparked from collective discussion of the Bernstein and Bourdieu synthetic theory base, were *explicit quality criteria*, *background knowledge*, *metalanguage*, and *narrative*. *Narrative* and *metalanguage* were developed, primarily by Allen Luke, as indicators of the specific pedagogical features identified as important in the critical literacy tradition. *Explicit quality criteria* was developed, primarily by Jennifer Gore, to pick up something lost in the collapsing of Bernstein's framing categories and linking it with well known calls for 'explicit teaching'. *Background Knowledge* synthesized Berlak and Berlak's focus on personal knowledge with the common call for teaching to build on students' extant cognitive schema, as well as the oft noted quasi-progressive notion of 'starting where the students are at.'

Entirely original to the QSRLS instrument were three items: *Inclusivity*, *Active Citizenship* and *Group Identities*. *Active Citizenship* was developed, primarily by Lingard, Mills and Hayes, as a means of identifying practices named in social education literature as necessary for the empowerment of all students. And *Group Identities*, also developed by Lingard, Mills and Hayes, was an attempt to articulate the positive identification of non-dominant social identities of students in the regular discourse of classrooms. The specification of group identities was derived from the philosophical work of Iris Young (xxxx). *Inclusivity* was built from a collective understanding of the ubiquitous slogan calling for inclusivity in classroom processes and curricular.

Overall then, the initial QSRLS observational instrument included 20 items, each cast in terms that allow rating their presence on a 1 to 5 ordinal scale. Since the development of the QSRLS observational instrument, substantial post-hoc theorizing has been done to help articulate the specific rationale for each item in the instrument, but the overall theoretical purposes need to be foregrounded here. From this overview it would be clear that one concern was to examine the unidimensionality of Authentic pedagogy more carefully. It was also intended to incorporate a broader estimation of the generalized notion of social support analyzed separately by CORS, into an interactional analysis of the effects of classroom practice. Finally, there was an explicit intent to include items in the measurement which the broader literature on educational inequality has long identified as potentially valuable for traditionally disadvantaged groups, but which had not been included in conventional effect studies of pedagogy. The overall strategy was to bring together previously divergent bodies of research into a methodology drawn from

conventional studies of school effects and effectiveness (a strategy articulated in Ladwig, 1996).

Here it is important to note the role this methodological strategy would play for each of the research agendas to which it was designed to respond. For socio-linguistic analysis of educational inequality, and for the radical sociology of school knowledge, the pay-off would be simple. This strategy provided a means by which to test some of the taken for granted theoretical claims in fields of research that focus on inequality but rarely, if ever, test the theoretical claims in a generalisable manner, with methods that are recognized and accepted outside their field.

For school effects and effectiveness research, the potential of the QSRLS classroom observational strategy was potentially even more compelling. As Dreeban and Gamoran (1986) pointed out about 20 years ago, in addition to the conventional organizational and curricular structure foci of the school effects tradition, it is necessary to take the nature and quality of instruction into model that seek to explain educational inequality. Indeed, the major trajectory of Gamoran's work throughout the 1990s was to do precisely that: model the quality of instruction within more structural analyses of tracking, precisely as he had declared (see Gamoran, 1989, where he points out that his previous work had primarily focused on quantity of content covered). Without taking anything at all away from the major contribution of Gamoran's work, it is necessary to point out, however, that the categories used to describe instruction were related to very general descriptive frameworks of classroom interactions as a discursive interchange, based on a general model of teaching (see Gamoran, et al, 1995). What these means is that these analyses essentially model generic qualities of teaching and find an unequal distribution of those qualities as a contributing factor in the resultant inequalities being modeled.

What this also means is that the categories used to model classroom interaction had little or nothing to do with theoretical explanations of educational inequality found in the more critical tradition in which Bernstein's and Bourdieu's conceptual frameworks are pivotal, known as the radical sociology of school knowledge (Ladwig 1986). If this the radical sociology of school knowledge is right, even if Gamoran were to find equally distributed qualities of teaching, as viewed through generic categories of 'good' teaching, equal student outcomes (by social group) would not be forth-coming. According to a more Bourdieuan framing, the cultural differences manifested in patterns of educational inequality are structured through internalized cognitive and dispositional differences that generic 'good' teaching literally could not reach. Like Gamoran (1989), though, the promise of the QSRLS modeling of pedagogy was to take the insights of qualitative and ethnographic studies that had generated this alternative theoretical framing, and put the theories to a test which might provide more defensible generalisable claims.

### **Modelling The First Wave of Data**

The first set of data was drawn from the first year's QSRLS field schools. For this sample of schools, eight were selected from the 104 *Leading Schools* cohort (in accordance with the contractual agreement). It is important to note that the selections of

schools was designed to be a stratified sample, covering urban and rural setting, high schools and primary schools, and both ends of the advantage-disadvantage spectrum, as defined by Education Queensland' measures of educational advantage (used to administer the Commonwealth equity programs). Apart from this stratification, however, the sampling was quite purposeful. That is, the project explicit set out to locate schools which were most likely to exhibit high levels of Authentic pedagogy. This decision was based the proposition than high levels of Authentic Pedagogy were not likely to be widely spread, and indeed rare. In order to maximize the possibility of having variance in our intended explanatory variables, such a sample was deemed essential.

Three grade levels had been selected to maximize comparability with the CORS SRS studies (which had focused on Years 6, 8, and 11) and at the same time maximize the likelihood of obtaining pre-existing prior achievement measures. In each of these year levels, four subject areas (English, Maths, Science and Social Sciences) were targeted as a primary sampling. In each school – year level combination, three classes were to be selected for observation (less in smaller schools where all classes were observed). As other portions of the QSRLS were more concerned with the school as an organizational unit, in addition to this specified sampling, each school in the study was also ask to nominate particularly interesting classes for us to observe and consider in our analyses of broader school features.

Observations in classrooms were conducted in two week-long field visits, scheduled to spread across the academic year. During each visit, each of the targeted classes was to be observed twice. At the end of this year of data collection, 302 classroom observation had been conducted. It was on this sample of 302 lesson observations that the first modeling was conducted. The now public formulation of Productive Pedagogy was the result of this (then tentative) modeling.

Between the second and third year of data collection, Queensland experienced a change of government and the abolition of the *Leading Schools* Project. The QSRLS continued on, however, as the initial findings had sparked some interest and the new government saw a need to complete the work. This also meant there was now the opportunity to sample schools from the entire state school system (not only the *Leading Schools*). At the time, this was seen, by the QSRLS team, as an important opening and opportunity to better search for the best pedagogy (in terms of the authentic model) we could find. By the end of the study, after three years of data collection, a total of 974 observations had been conducted from 24 schools. For the purposes of this analysis both I will first discuss first year modeling before discussion full sample analyses.

### **Year One Modeling Results**

Before turning to the modeling behind Productive Pedagogy, I will present the descriptive findings needed to discuss the more analytical results. Table one, below, presents the final sample distribution across subject area and year levels.

#### **Table One: Final Observation Sample, Year Level by KLA**

	Curriculum Area					Total
	Math	English	Social Sciences	Science	Other	
<b>Year 6</b>	<b>73</b>	<b>66</b>	<b>37</b>	<b>42</b>	<b>47</b>	<b>265</b>
<b>Year 8</b>	<b>65</b>	<b>59</b>	<b>64</b>	<b>68</b>	<b>46</b>	<b>302</b>
<b>Year 11</b>	<b>69</b>	<b>68</b>	<b>67</b>	<b>86</b>	<b>10</b>	<b>300</b>
<b>Other</b>	<b>18</b>	<b>36</b>	<b>15</b>	<b>5</b>	<b>33</b>	<b>107</b>
<b>TOTAL</b>	<b>225</b>	<b>229</b>	<b>183</b>	<b>201</b>	<b>136</b>	<b>974</b>

As can be noted from Table One, approximately 14% of the observations were outside of the main four KLAs, and around 11% were conducted outside of the target years. This included virtually every subject area in Queensland and all year levels from pre-school setting to adult learning settings (which were part of sample school special programs). Across the three years of the study, roughly equal numbers of observations were conducted each year (302 in 1998, 343 in 1999, 329 in 2000).

Interestingly, although the opening up of the school sampling to be selected from the all government schools in Queensland was associated with a steady yearly increase in the QSRLS classroom observational item ratings, the overall pattern of these ratings did not alter substantially. Table Two, below, presents the descriptive statistics for each individual item rating for the full QSRLS sample (974 observations). From this table a number of patterns can be noted that were true for every year of the study (meaning for every set of 8 schools). Keeping in mind the rating scales were from 1 to 5, it is notable that very, very few of the items had a mean above the theoretical mean of 3. That is, in general terms the ratings were low. *Students' self-regulation*, *Inclusivity*, and *Social Support* were the exceptions to this pattern, throughout the study. Similarly consistent throughout the study was the fact that three items were so notably low that their measures of skewness (the degree to which a frequency distribution is crammed on one side of the scale) and Kurtosis (the flatness of a frequency distribution) were so high that including them in models is statistically problematic (since they essentially didn't vary, see Appendix A for a graphic represent of this).

**Table Two: Descriptive Statistics of Elements in the QSRLS Observational Instrument, Full Sample**

	Mean	Standard Error	Standard Deviation	Skew	Kurtosis
<b>Students' Direction</b>	1.59	0.03	0.85	1.48	1.97
<b>Knowledge Integration</b>	1.76	0.04	1.10	1.57	1.72
<b>Problematic Knowledge</b>	1.74	0.03	1.06	1.41	1.14
<b>Cultural Knowledges</b>	1.35	0.03	0.79	2.56	6.38
<b>Higher Order Thinking</b>	2.55	0.04	1.11	0.16	-0.85
<b>Deep Knowledge</b>	2.71	0.03	1.07	0.12	-0.65
<b>Deep Understanding</b>	2.60	0.03	1.02	0.13	-0.47
<b>Substantive Conversation</b>	2.27	0.04	1.20	0.62	-0.60
<b>Social Support</b>	3.66	0.03	0.96	-0.18	-0.67
<b>Engagement</b>	3.73	0.03	1.01	-0.40	-0.57
<b>Inclusivity</b>	4.04	0.04	1.19	-0.99	-0.16
<b>Background Knowledge</b>	2.61	0.04	1.17	0.50	-0.57

<b>Connectedness</b>	1.90	0.04	1.09	1.04	0.14
<b>Explicit Criteria</b>	2.13	0.04	1.10	0.88	0.14
<b>Self-Regulation (Student)</b>	4.03	0.03	0.97	-0.83	0.10
<b>Problem-Based Curriculum</b>	2.01	0.04	1.30	1.06	-0.12
<b>Narrative</b>	1.95	0.04	1.13	1.13	0.43
<b>Group Identities</b>	1.19	0.02	0.62	3.81	15.39
<b>Citizenship</b>	1.18	0.02	0.57	3.58	13.49
<b>Metalanguage</b>	1.75	0.03	1.02	1.43	1.50
<b>NB: n = 974</b>					

**Table TwoA: Descriptive Statistics of Elements in the QSRLS Observational Instrument, Year One Sample**

	<b>Mean</b>	<b>Standard Error</b>	<b>Standard Deviation</b>	<b>Skew</b>	<b>Kurtosis</b>
<b>Students' Direction</b>	1.36	0.04	0.65	1.89	3.34
<b>Knowledge Integration</b>	1.61	0.06	0.97	1.83	2.90
<b>Problematic Knowledge</b>	1.61	0.06	1.00	1.74	2.33
<b>Cultural Knowledges</b>	1.25	0.04	0.63	2.89	8.90
<b>Higher Order Thinking</b>	2.40	0.06	1.12	0.30	-0.87
<b>Deep Knowledge</b>	2.57	0.06	1.05	0.18	-0.78
<b>Deep Understanding</b>	2.43	0.05	0.93	0.20	-0.37
<b>Substantive Conversation</b>	2.16	0.07	1.17	0.65	-0.66
<b>Social Support</b>	3.31	0.05	0.89	-0.08	-0.39
<b>Engagement</b>	3.34	0.06	1.00	-0.19	-0.56
<b>Inclusivity</b>	3.55	0.07	1.29	-0.45	-0.90
<b>Background Knowledge</b>	2.32	0.07	1.17	0.71	-0.37
<b>Connectedness</b>	1.64	0.05	0.95	1.32	0.67
<b>Explicit Criteria</b>	2.08	0.06	1.07	0.90	0.21
<b>Self-Regulation (Student)</b>	3.69	0.06	1.05	-0.43	-0.58
<b>Problem-Based Curriculum</b>	1.81	0.07	1.15	1.27	0.53
<b>Narrative</b>	1.90	0.07	1.15	1.27	0.74
<b>Group Identities</b>	1.14	0.03	0.52	4.45	21.43
<b>Citizenship</b>	1.16	0.03	0.51	3.71	14.58
<b>Metalanguage</b>	1.77	0.06	1.04	1.26	0.81
<b>NB: n = 302</b>					

After the first year of the study, with the first 302 observations in hand, initial modeling of the data was conducted to test if these data fit theoretical models of the underlying dimensions, or factors, of classroom practice the QSRLS was seeking to examine. At the point, the QSRLS had more than one working model of classroom practice, and the task was to find which model fit the data best.

At the same time, consistent with the continual adjudication between theory and evidence in modeling, a four dimensional model was the preferred theoretical choice as it was designed to allow the examination of our larger theoretical concerns. As implied above, in essence the QSRLS had intended to model varying interactions of dimensions of pedagogy to see 1) if the general findings of CORS would obtain, and 2) to examine the effects of these dimensions relative to students from traditional recognized equity groups, in particular students from low socio-economic, Indigenous, non-English speaking

backgrounds and both sexes. That is, in addition to the authentic pedagogy and the social support dimension identified by CORS, the QSRLS was particularly interested in modeling those pedagogical practices identified by equity literature as desirable.

The modeling of the first years' data was initially completed with congeneric confirmatory factor analysis procedures (using LISREL). The first task in this modeling was to examine the Authentic Pedagogy construct itself, using the disaggregation of Newmann's depth item (with their error measures of the two new depth items co-varied). The results of this first analysis confirmed doubts about Authentic Pedagogy as a uni-dimensional construct.

Two estimates from the CFA modeling of Authentic Pedagogy suggested it was not uni-dimensional. First, the individual contribution of the variable 'connectedness to the world beyond the classroom' (0.46) was approximately half that of the other variables (whose loadings ranged from .80 to .89), indicating its relative weak contribution to the underlying factor, Authentic Pedagogy. Second the error term for 'connectedness to the world' was quite large (.79), suggesting that the major of that particular item was measuring something else.<sup>6</sup> This finding supported the QSRLS initial arguments about the need to understand classroom practice as multi-dimensional in contrast to Newmann's uni-dimensional concept and the expansion of the SRLS classroom observation scale to include a number of other measurement items.

Removing the 'connectedness' item from the Authentic Pedagogy construct substantially narrows its focus, in a manner similar to Newmann's own work after CORS, in which questions of knowledge construct, disciplined inquiry and substantive communication clearly become more an issue of understanding the nature and processes through which students' experiences can be said to represent intellectual quality (see Newmann, Bryk, Nagata XXXX). Following this line of reasoning, the next modeling examined intellectual quality specifically. The remaining items from Newmann's AP construct (higher order thinking, the two depths, substantive communication) plus two new QSRLS items (*problematic knowledge* and *metalinguage*) were examined as indicators of Intellectual Quality. Overall this model was quite stable. Allowing error estimates of higher thinking to vary with the depth of knowledge (which is conceptually readily defensible), the relative contribution of all items was substantial (although metalinguage was weakest at .51), and the overall fit was solid.<sup>7</sup>

Modelling *connectedness to the world beyond the classroom* as an indicator of an underlying measure of the degree to which classroom interaction was connected into some context beyond the immediate lesson also proved a relatively straightforward matter. Taking *connectedness*, *background knowledge*, *knowledge integrations* (links between subject areas), and *problem-based curriculum* all as indicators of underlying degree of contextualization present in a lesson, these four items were modeled as measures of broader dimensions, later labelled Connectedness. Individual item contributions ranged fairly evenly from .62 to .72, and overall fit measures were quite defensible.<sup>8</sup>

The social support indicators from CORS also worked well in a broader modeling of the observational items related to the degree to which the classroom could be characterized as having a socially supportive environment. Included in the measurement of this construct were *problem-based curriculum*, *student control*, *self-regulation* and *explicit quality criteria* added to the CORS items, *social support* and *engagement*. Similar to the Connectedness modeling, contributions amongst the items were relatively comparable (in the .7 range) with the exceptions of students' *self-regulation* and *explicit quality criteria* which behaved as a pair with lower contributions (.51 and .53 respectively) and co-varying error estimates. The overall fit of socially supportive environment model was also readily defensible.<sup>9</sup>

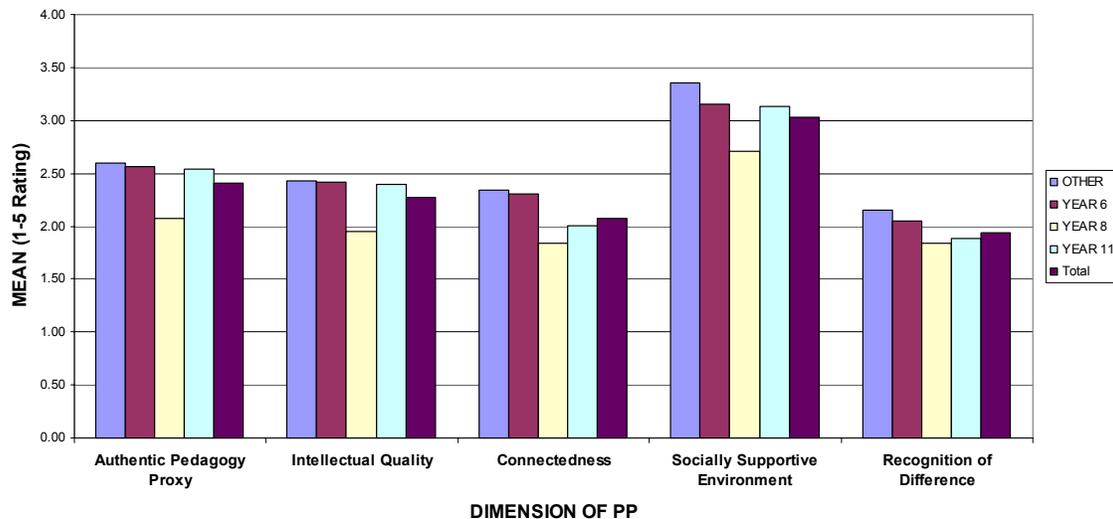
From this point the modeling exercised was less smooth. Recall that the remaining items had been constructed to estimate an underlying dimension of pedagogy related to what theoretically should assist students from disadvantaged background. Since three of the remaining items were strongly lacking in variance (*cultural knowledge*, *active citizenship* and *group identifies*) and obtained very high measures of skew and kurtosis, these items are not at all suitable for CFA estimations<sup>10</sup>. However, since the opening of the potential pool from which school would be selected had just occurred, the possibility of finding more sites where these items might be observed to a greater extent existed. Consequently, as the preferred theoretical model included a fourth underlying factor, initial estimates of this factor were estimated as a tentative iteration in the overall project. Individual item contributions varied from .40 to .72, but the overall fit was substantially less than optimal, a situation readily attributable to the lack of variance in the constituent items.<sup>11</sup>

### **What did these constructs show about classrooms? Descriptive Results**

With three constructs working well, and the fourth seemingly (at that point) recoverable, these four constructs were reported to Education Queensland as four dimensions of an overall model of classroom practice then tentatively named *Productive Pedagogy*. Even with their weakness in technical measurement terms, these constructs potentially still held theoretical and descriptive validity for the QSRLS. Proceeding on this tentative basis, the first simple task of the QSRLS was simply to report the relative frequency of classroom practices summarized by the four dimensions of Productive Pedagogy. To facilitate this in a manner that allowed straight forward conversion to our coding categories, summative scales for each dimension were calculated (as mean scores on the 1 to 5 scale) and reported for each of the three main sample year-levels (year 6, 8 and 11).

Figure One, below, presents a column chart of these descriptive mean scores for each of the dimensions of Productive Pedagogy, broken into five categories: Other, Year 6, Year 8, Year 11 and full sample. The 'other' category included teachers identified by the school as exemplary of their intended or desired restructuring. While this simple comparative presentation was first done with only the observations from the first year of the study, Figure One presents the final full sample from all three years (total N = 974 observations). The reasons for presenting full sample results here are two-fold. First, despite the hopes of the QSRLS team at the end of the first year, the overall pattern

of observation results only changed slightly, with only slight overall increases each year of the study for each dimension mean, (which only became statistically significant once the sample of observations went over 800). Thus the general pattern presented in Figure One held true each year of the study. Second, the overall pattern of results was one of the findings from the QSRLS that help policy significance in the reception of the QSRLS. The patterns identified in Figure One were one of the reasons the dimensions identified in the initial modeling had become part of Education Queensland teacher development initiatives before the QSRLS was completed.



**Figure 1: Mean Ratings Productive Pedagogy Dimensions, by Year-Level, Full Sample N=974**

There are four notable findings presented in Figure 1. First, it is clear that over the three main sample year-levels, Year 8 was substantially lower in all four dimension than either year 6 or year 11. (It is important to keep in mind that Year 8 is the first year of secondary school in Queensland.) This difference between year levels was statistically significant at every point in the study (at the  $p < .01$  level and beyond as the sample grew). Second, the fact teachers identified by the schools themselves scores higher than the teacher who were selected as part of the stratified sampling procedures suggests that what is identified by Productive Pedagogy at least corresponded to what the schools themselves identified as desirable teaching. Third, the Socially Supportive Environment Dimension consistently averaged higher means across all year levels. And fourth, the other three dimension means, *Intellectual Quality*, *Connectedness* and *Recognition of Difference* did not even come close to the theoretical mean of 3.

While the rubrics through which these dimensions were never calibrated so as to allow direct comparison, linking the mean item scores within the dimensions to the rubric categories for those mean scores does provide a rough qualitative sense of just the mean scores indicate. For example, high order thinking and depth were occasionally and unevenly part of the typical lesson, while metalanguage, substantive conversations and problematic knowledge were rare. Thus the overall mean score for Intellectual Quality

reflects a generally low level of intellectual demand in classrooms. In the other hand, most students, most of the time, were on-task, reasonably self-regulated and engaged. At the same time, it was very rare for students to have control over deciding the direction of their lesson activities and equally rare that they were presented with criteria by which the quality of their work would be judged.<sup>12</sup> In the other dimensions, most aspects of Connectedness were rare and as noted above, several items in the *Recognition of Difference* scale were essentially non-existent.

For a variety of reasons, it was from this basic level of descriptive reporting, first reported in the March 1999 QSRLS report to Education Queensland, that the Productive Pedagogy model very rapidly became widely distributed, publicized and indeed incorporated into the professional development lexicon of Education Queensland (well before the development of the New Basics initiative). Irrespective of any measurement concerns, or even any demonstrable effect of the dimensions within Productive Pedagogy, this model became popular among teachers and school leaders as reflective tools for teachers to individually and collectively analyse current practice.<sup>13</sup> Indeed, so popular was the model that school and principal interest in it within New South Wales led to the development of the NSW model of Quality Teaching, from Ladwig and Gore.

## Conclusions

From the initial descriptive modeling of pedagogy within the QSRLS, two main conclusions can be safely drawn. First, from a substantive point of view, it is evident that the initial notion that the *Authentic Pedagogy* construct really should not be taken to be a singular, unified reality in classroom practice has been substantiated by the QSRLS classroom observational data. To demonstrate this further, a Principal Components confirmatory factor analysis was conducted based on each of the items from the Authentic Pedagogy construct along with those items of the QSRLS observation instrument which are theoretically most linked to the contextualization implied by the CORS concern for value beyond school. The results of this analysis are present below in Table Three.

Table Three: Rotated Component Matrix for Authentic Pedagogy and Connectedness

	<b>Component 1</b>	<b>Component 2</b>
<b>Higher Order Thinking</b>	<b>0.84</b>	0.26
<b>Depth of Knowledge</b>	<b>0.87</b>	0.22
<b>Depth of Understanding</b>	<b>0.88</b>	0.17
<b>Substantive Communication</b>	<b>0.70</b>	0.37
<b>Connectedness</b>	0.19	<b>0.77</b>
<b>Background Knowledge</b>	0.27	<b>0.74</b>
<b>Knowledge Integration</b>	0.14	<b>0.75</b>
<b>Problem-Based Curriculum</b>	0.43	<b>0.54</b>
<b>Explained Variances</b>	53%	14%

As is evident in Table Three, the original CORS item ‘connected to the world beyond the classroom’ loads onto a second factor identified by this bank of items – with the first factor indicative of the QSRLS dimension, *Intellectual Quality*, and the second component being the QSRLS *Connectedness* dimension. On the basis of this and the congeneric CFA analysis, whatever the other results of Productive Pedagogy, this evidence clearly suggests that the singularity of Authentic Pedagogy needs to be remain open to further research.

The second main conclusion to be drawn from the initial modeling done in the development of Productive Pedagogy is that the extant construct of Recognition of Difference can not be defended as a reliable empirical phenomenon. That is, however much educators might well believe in, and wish to defend, the importance of recognition of difference as a concept, the measures within Productive Pedagogy do not provide any defensible empirical basis for such advocacy. To demonstrate this further beyond the distributional problems of individual items in the Recognition of Difference scale, consider the measures of inter-item reliability (or internal consistency) for each of the dimensions measured by the Productive Pedagogy instrument, estimate with Cronbach’s Alpha, presented below in Table Four.

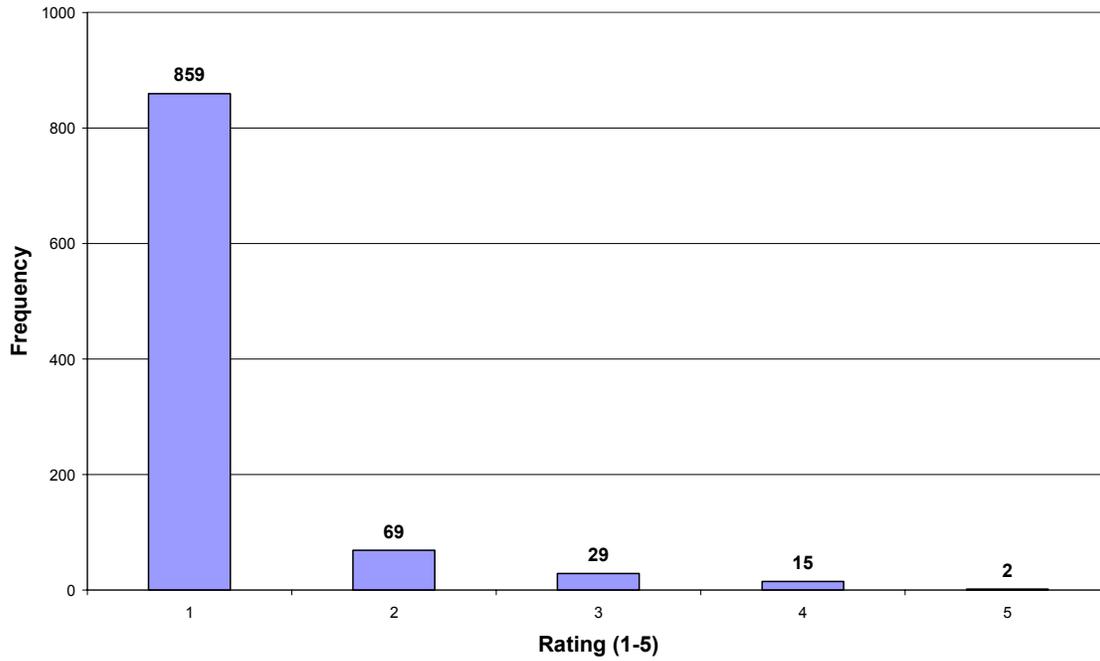
Table Four: Cronbach Alpha’s estimates for PP Dimensions

<b>Internal Consistency Reliability, N=974, full sample</b>	<b>Cronbach Alpha</b>
Authentic Pedagogy Proxy	.85
Intellectual Quality	.85
Connectedness	.74
Supportive Classroom Environment	.73
Recognition of Difference	.53

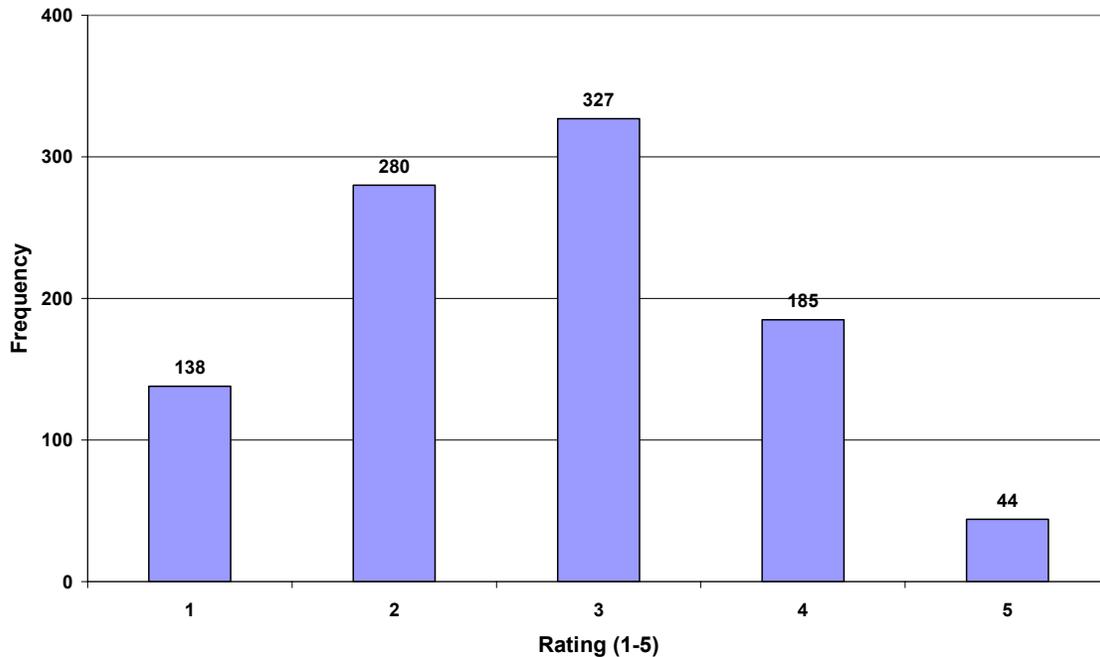
While there may be some differences amongst statisticians as to what range of a alpha is acceptable, I know of no standing defense for an alpha as low as that obtained by the Recognition of Difference scale. Clearly this scale is open to classical Type II error inferences. In lay terms, there are two possible reasons for the skewed distribution of items in Recognition of Difference that led to such a low internal consistency measure. One the one hand, it is possible, for example, that the Active Citizenship measure looks the way it does simply because very few teachers ever actually do anything recognizable as citizenship. On the other hand, it is equally possible that the item Active Citizenship was just poorly defined and therefore didn’t measure what citizenship did occur in the classrooms observed. The QSRLS data will not provide reliable evidence to decide which of these possible interpretations is a better one. Consequently, while it may be the case the Recognition of Difference exists and is valuable, it is certainly not that case the QSRLS has a reliable measure of just what this is.

## Appendix A

The following bar graphs give a graphic sense of the frequency distributions discussed above.



**Figure 2: Active Citizenship Ratings, Frequencies**



**Figure 3: Deep Knowledge Ratings, Frequencies**

## REFERENCES

- Berlak, A. and Berlak, H. (1981) *Dilemmas of Schooling*. London: Methuen.
- Dreeban, R and Gamoran, A. (1986). Race, Instruction, and Learning. *American Sociological Review*, Vol. 51, No. 5, 660-669.
- Gamoran, A. (1987a). The Stratification of High School Learning Opportunities. *Sociology of Education*, Vol. 60, No. 3, 135-155.
- Gamoran, A. (1987b). Organization, Instruction, and the Effects of Ability Grouping: Comment on Slavin's "Best-Evidence Synthesis." *Review of Educational Research*, Vol. 57, No. 3, 341-345.
- Gamoran, A. (1989). Measuring Curriculum Differentiation. *American Journal of Education*, Vol. 97, No. 2, 129-143.
- Gamoran, A. (1992a). Access to Excellence: Assignment to Honors English Classes in the Transition from Middle to High School *Educational Evaluation and Policy Analysis*, Vol. 14, No. 3, 185-204.
- Gamoran, A. (1992b). The Variable Effects of High School Tracking. *American Sociological Review*, Vol. 57, No. 6, 812-828.
- Gamoran, A. (1993). Alternative Uses of Ability Grouping in Secondary Schools: Can We Bring High-Quality Instruction to Low-Ability Classes? *American Journal of Education*, Vol. 102, No. 1, 1-22.
- Gamoran, A, and Berends, M. (1987). The Effects of Stratification in Secondary Schools: Synthesis of Survey and Ethnographic Research. *Review of Educational Research*, Vol. 57, No. 4, 415-435.

- Gamoran, A. Nystrand, M. Berends, M., and P. C. LePore. (1995). An Organizational Analysis of the Effects of Ability Grouping. *American Educational Research Journal*, Vol. 32, No. 4, 687-715.
- Gamoran, A., Porter, A., Smithson, J, and P. A. White. (1997). Upgrading High School Mathematics Instruction: Improving Learning Opportunities for Low-Achieving, Low-Income Youth. *Educational Evaluation and Policy Analysis*, Vol. 19, No. 4, 325-338.
- Ladwig, J. G. 1995. Curriculum and Social Identity. Paper presented at the annual Sociology of Education conference, Sheffield, England. (January)
- Ladwig, J.G. 1998. Authentic school reform. Discourse.
- 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/>)

---

## ENDNOTES

<sup>1</sup> I should note that the plural form of this label, Productive Pedagogies, has also be used publicly to denote ostensible variability in desirable forms of classroom practice. I maintain the singular in this analysis since, as will be shown, the research on which the label is based does not in itself support the denotation of the plural form. Additionally, to my knowledge, formally 'pedagogy' has no plural form.

<sup>2</sup> For a more skeptical view of Education Queensland's intent and a more elaborated historical account of these events, see Lingard XXX. There is an irony in the way Education Queensland wedded Authentic Pedagogy to SBM, in that the work done by CORS actually had demonstrated that SBM was un-related to improvements in pedagogy and student performance, especially in terms of Authentic Pedagogy (see, Newmann, 1993).

<sup>3</sup> Thus the initial design of the QSRLS was largely a replication and extension of core studies in Newmann's centre. On the one hand, detailed, multi-method data was to be gathered from a total of 24 schools over three years – 8 schools per year. While survey data was gathered from all students in target year-levels and all teachers in these schools, the proposed 24 schools were the sites of direct classroom observations, interviews with teachers, school leaders and key 'change agents', and the source of samples of student work samples and teacher assessment tasks. Taken together, these schools became known as the QSRLS field study schools. In addition to this set of data, a broader survey of 150 schools was to be conducted annually, in a cascading longitudinal design over the three year life of the study. This second set of data was known simply as the annual survey. Finally, as a third set of data, Education Queensland committed itself to providing student-level prior achievement data and school level demographic and student performance data. All of these data sets were to be brought together in a multi-level analysis of school restructuring in a manner similar to CORS, covering four levels of phenomenon: student performance and experience, classroom practices, school organizational capacity, and external 'supports' (cf Newmann and Wehlage, 1995).

While this design was intended as a direct response to the Queensland interest in Authentic Pedagogy and SBM, it was also helpful that the designer and author of the initial proposal (and of this analysis) was a member of the CORS team that had developed the instruments used to Authentic Pedagogy and school restructuring, and had been part of the conduct of the CORS studies prior to emigrating to Australia. Thus, access to all levels of CORS information (including the unreported material) was direct for the QSRLS.

<sup>4</sup> It should also be noted that many educational commentators have suggested that the academic focus of authentic pedagogy was itself a problem, arguing that 'some' students just don't benefit from an academic concentration. Without opening the debate about the normative ends of schooling, I would just note that this view clearly has not taken into account the long recognized positive impact of academic press for students from all socio-economic and racial backgrounds in the us (See Coleman, et al; Lee xxx)

---

<sup>5</sup> If this line of reasoning isn't clear, it may help to consider the simple proposition that there are many ways a teacher might produce classroom practices that would score highly on the authentic pedagogy measure. Some of these differences were documented in the reports on authentic pedagogy, others were known first hand to the author, who had conducted some of the field work on which CORS analyses were based.

<sup>6</sup> The goodness of fit indexes for the CFA model of Authentic Pedagogy also yielded mixed results. While overall measures were quite good (GFI=.99, AGFI=.99, etc.) the Chi-square and Root Mean Square Error of Approximation were less than optimal (.04 and .07 respectively). [PUT IN REFERENCE ON FIT RULES OF THUMB TO BE USED]

<sup>7</sup> The GFI=.99 and AGFI=.98, while the Chi-square p was .09 and RMSEA=.05.

<sup>8</sup> GFI and AGFI = .99, Chi-square p= .25, and RMSEA = .04.

<sup>9</sup> GFI = .99, AGFI = .97, Chi-square p= .29, and RMSEA = .03.

<sup>10</sup> One generous rule of thumb holds 3 as an upper boundary for skew and two of these items were way beyond that. More conventional rules of thumb hold 2 as a cut-off and all three items were beyond that threshold of acceptability.

<sup>11</sup> GFI = .94, AGFI = .83, Chi-square p = .00, and RMSEA = .06.

<sup>12</sup> To follow the concordance from the item means to the rubric descriptors, public access to a large portion of the rubrics can be found through Education Queensland:  
[http://education.qld.gov.au/public\\_media/reports/curriculum-framework/productive-pedagogies/html/manual.html](http://education.qld.gov.au/public_media/reports/curriculum-framework/productive-pedagogies/html/manual.html). In that public version, the codes for a 1, a 3, and a 5 were included. A full version of the Productive Pedagogy observation manual has also been distributed widely in Queensland and NSW, but never published. Full copies can be obtained from the Professional Learning in Education Research Group, School of Education, The University of Newcastle, Callaghan, NSW, 2308.

<sup>13</sup> Another measurement concern not noted above included inter-rater reliabilities. To monitor this, approximately 20% of the lessons observed were independently coded by two observers and their scores analysed for 1) the percentage of exact agreement and the percent of agreement within one rating, and 2) the Pearson correlation coefficient. In general, the items of the first three dimensions, Intellectual Quality, Connectedness and Socially Supportive Classroom Environment obtained levels of inter-rater reliability comparable to those reported by Newmann for Authentic Instruction (see Newmann and Associates, 1996, appendix A). That is, within one agreement occurred generally around 90% of the time (some items higher, none lower than 85%), and the correlations were generally above .65 (although strongly skewed items had more exact agreement and lower correlations due to than skew). Three items did not obtain these levels of inter-rater reliability, metalanguage, narrative, and inclusivity. The specification of these items' rubrics was slightly refined after the first year of the study to improve reliability. By the end of the study, the reliability of metalanguage and narrative were up to the levels of the other items. Inclusivity never obtained highly defensible levels of reliability.