Quality Assessment in University Social Science Courses

Jennifer Gore, James Ladwig, Wendy Amosa, Tom Griffiths, Robert Parkes and Hywel Ellis
The University of Newcastle

Paper prepared for presentation at the Australian Association for Research in Education annual meeting, Brisbane,

30 November - 4 December, 2008.

While substantial energy has gone into understanding the mechanics of assessment in higher education, little attention has been paid to developing means by which university lecturers can monitor the quality of the assessment tasks they develop. This paper introduces a research project, funded by the Australian Learning and Teaching Council (formerly the Carrick Institute), which was designed to: (a) enhance the quality of assessment in the social sciences in the tertiary sector; and (b) refine and evaluate a model for analysing and improving the quality of assessment tasks in the social sciences, primarily in first year courses. The investigation draws on a model for high quality assessment practice, known as the Quality Teaching model (see Table 1), initially developed to enhance practice in school education (NSW Department of Education and Training, 2004). The model, which is comprised of three dimensions – Intellectual Quality, Quality Learning Environment, and Significance – has now been implemented in schools across NSW and has been adopted by other states and territories in Australia, as well as overseas, as part of a broader move to improve teaching.

Intellectual Quality	Quality Learning Environment	Significance
DK - Deep knowledge DU - Deep understanding PK - Problematic knowledge HOT - Higher order thinking M - Metalanguage SC - Substantive communication	EQC - Explicit quality criteria HE - High expectations SD - Student direction	BK - Background knowledge CK - Cultural knowledge KI - Knowledge integration C - Connectedness N - Narrative

Table 1. Dimensions and elements of the Quality Teaching model

Substantial applied research has been conducted in relation to Quality Teaching in primary and secondary school contexts with demonstrated improvements in both outcomes and equity (Amosa, Ladwig, Griffiths, & Gore, 2007; Ladwig, Smith, Gore, Amosa, & Griffiths, 2007). However, there has been little implementation of these ideas in tertiary education contexts to date. In exploring implications of Quality Teaching at the University level, this study potentially makes an empirical contribution to contemporary debates on the distinction between pedagogy and andragogy (Knowles, Holton, & Swanson, 1998).

The major focus of the study is on the link between task quality and student performance. In general, assessment tasks are presumed to provide a reasonable basis for judging whether students have met the outcomes of a course and to differentiate among students. Common complaints from students about the design of assessment tasks relate to: misalignment with the stated curricular goals; misalignment with the teaching that has been provided; lack of clarity about expectations; unreasonably high or low expectations; and, tasks that hold little significance for students' career or professional aspirations (Boud & Falchikov, 2006; Knight, 2002; Shavelson, 2007).

This project differs from existing efforts to articulate standards for university teaching in its focus on articulating the underlying principles by which assessment practices can be evaluated and improved. Familiar measures of effectiveness in tertiary education centre on: retention rates, student progress through coursework, student engagement, and students' ratings of teaching or course quality. In our

Support for this publication has been provided by The Australian Learning and Teaching Council Ltd, an initiative of the Australian Government Department of Education, Employment and Workplace Relations. The views expressed in this publication do not necessarily reflect the views of The Australian Learning and Teaching Council.

study, the focus is on naming what counts as high quality assessment. Beginning with exploring conceptual and procedural issues in the refinement of assessment tasks is a practical avenue for opening what is ultimately part of a much larger foray into improving the quality of teaching in universities.

Aims

The project has the following key aims:

1. to evaluate the applicability and specify the components of a model for analysing the quality of assessment tasks in the tertiary sector;
2. to use the model to provide an 'audit' of the quality of assessment tasks across disciplines in the social sciences as a basis for collaborative task refinement; and
3. to examine the correlation between the assessment tasks and student outcomes.

Methodology

Academics from five disciplines from one university are participating in the research by providing for analysis: (a) assessment tasks being implemented in social science courses; and (b) student work samples produced in response to those assessment tasks. The disciplines represented are from the faculties of: Education (courses include; educational psychology, pedagogy, social contexts, and curriculum studies); Law; Humanities and Social Sciences (courses include: social work, language studies, speech pathology, history and sociology); Medicine and Public Heath; and Nursing and Midwifery (courses include: Concepts of health and wellness and mental health literacy and mental health first aid).

The overall design of the project involves two stages. Stage one is an initial 'audit 'of assessment tasks, including measures of the corresponding student work (described below) followed by professional learning about the model with the participating academics, both in relation to the tasks they have submitted for the project and more generally in relation to the model. Discussions with the participating academics will form the basis for refinement of the model of assessment practice. Stage two involves implementation of the refined model and repeat of the measures of task quality and corresponding student work. Another workshop with participating academics will be held to further refine the model prior to wider dissemination of the model in the tertiary sector.

The 'audit' of tasks in the social sciences has been conducted. Also, de-identified samples of randomly selected student work have been analysed on the following criteria: disciplinary depth, level of analysis, richness of written work, and treatment of knowledge. The instrument used for this analysis was adapted from the Authentic Achievement scales which have been shown to be strong predictors of (school) student performance on conventional standardised tests (Newmann, 1996). Moreover, these scales for analysing student work would appear to be consistent with the emphasis on intellectual depth one might expect in universities.

Findings

This paper reports only preliminary findings relating to a portion of the stage one data from the project. That is, of the total amount of student work collected for stage one, approximately 40% has been coded, while approximately 60% of the assessment tasks have been coded.

Task Quality

The quality of the assessment tasks, as coded on a 1-5 scale using the Quality Teaching Assessment instrument, where 5 is high and 1 is low, is depicted for each dimension of the model in Figures 1-3. These data show considerable variance in average scores for the different elements of the model with Deep Knowledge, for example, coding much higher than Problematic Knowledge. This variability in scores for particular elements of the model is consistent with ratings of secondary and primary school assessment tasks, while the overall ratings for most elements are higher than we found in either secondary school assessment tasks across a range of subject areas or in secondary school assessment tasks for the subject Human Society and its Environments (HSIE) which is the closest school subject to the social sciences focus on our study (Figures 4-6).

Some results are of particular interest in a university context include the following. In general, the assignments appear to be intellectually challenging, and in some cases provide high levels of student control over the substance of their responses with choice encouraged (Student Direction). On the other hand, while the problematic nature of disciplinary knowledge is addressed and students are, to some degree, required by the assessment tasks to explore the construction of knowledge in their responses, the level of Problematic Knowledge is not high. In a University context, in which complexity and disciplinary depth might be expected, higher scores for this element were anticipated.

Furthermore, given that university courses typically play a significant role in introducing students into specific discourse communities, the element of Metalanguage was surprisingly low with a mean of 1.29. Explicit Quality Criteria, while much higher scoring with a mean of 3.0, was coded with the full range of 1-5 for this set of tasks. This result means that there are many instances in these courses in which students are given no criteria as to what constitutes good work for a particular task. Without explicit criteria, students can expend considerable energy trying to decipher what is required or trying to guess what is in the lecturer's head.

Tasks requiring students to recognize and value non-dominant cultural knowledge (CK) or to integrate knowledge across subject areas (KI) were particularly rare, with mean scores of 1.71 for each. The importance understanding non-dominant cultural knowledge, particularly in fields of professional practice, may require more attention to this element in the design of assessment tasks. Also, given the number of courses in the study that are part of professional preparation programs, a higher level of integration may be desirable in assisting students to develop a coherent body of conceptual and practical knowledge with which to guide their developing professional repertoires and identities.

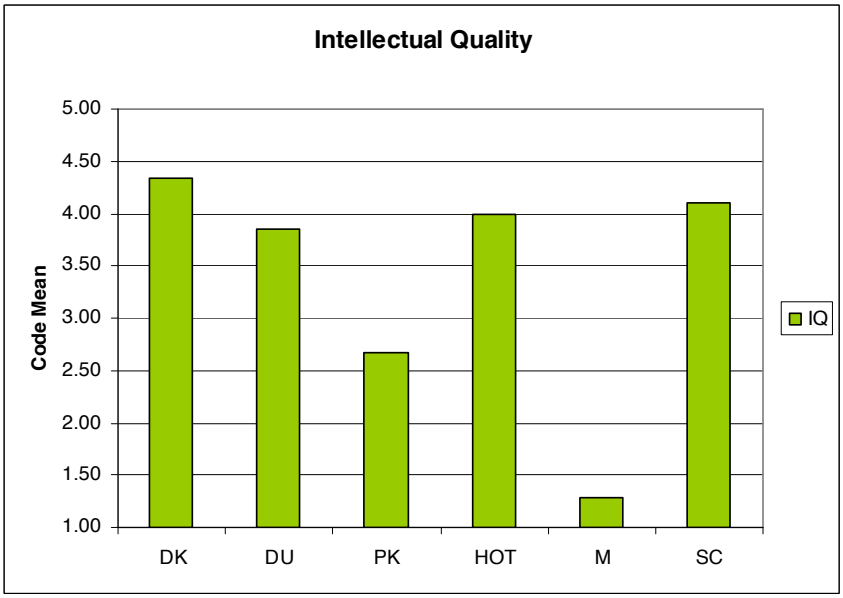

Figure 1. Task Intellectual Quality scores (n=22)

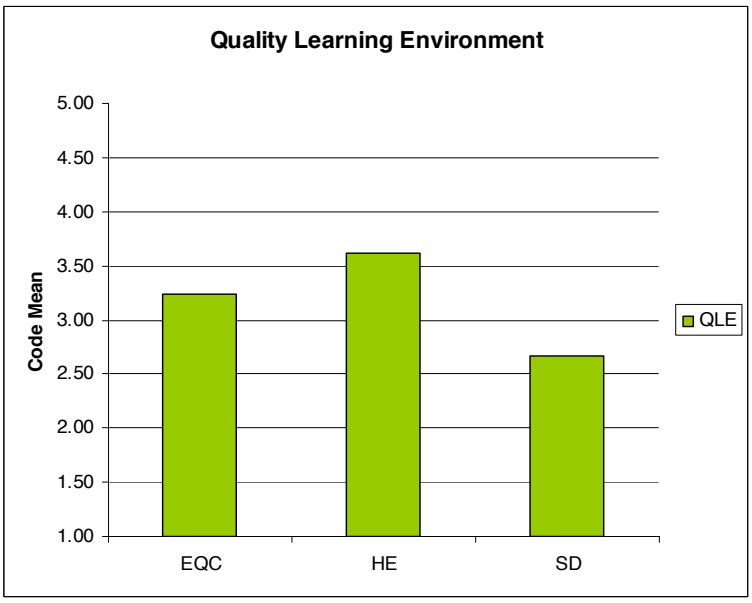

Figure 2. Task Quality Learning Environment scores (n=22)

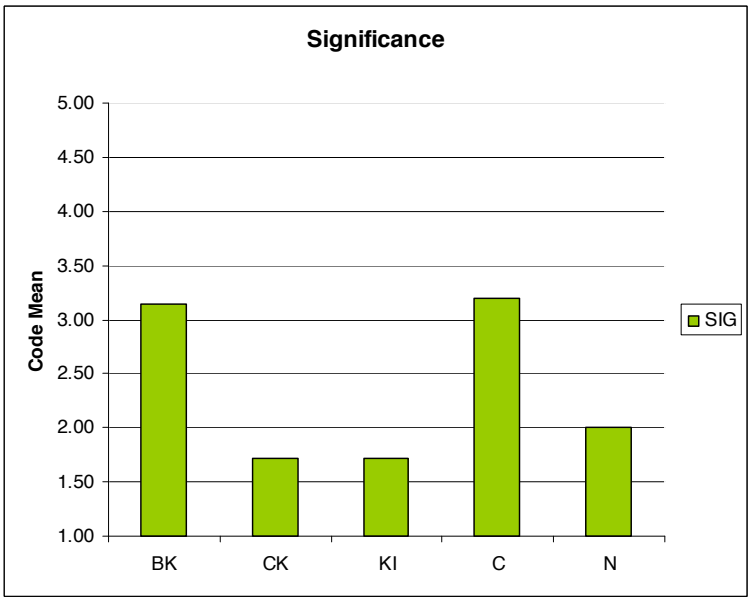

Figure 3. Task Significance scores (n=22)

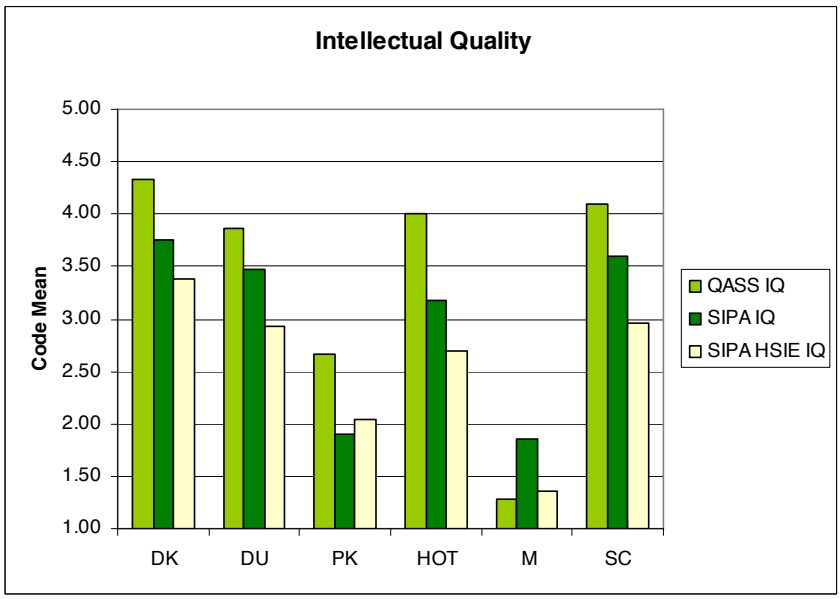

Figure 4. Comparison of task Intellectual Quality scores among three samples: University social science courses, secondary courses, HSIE courses

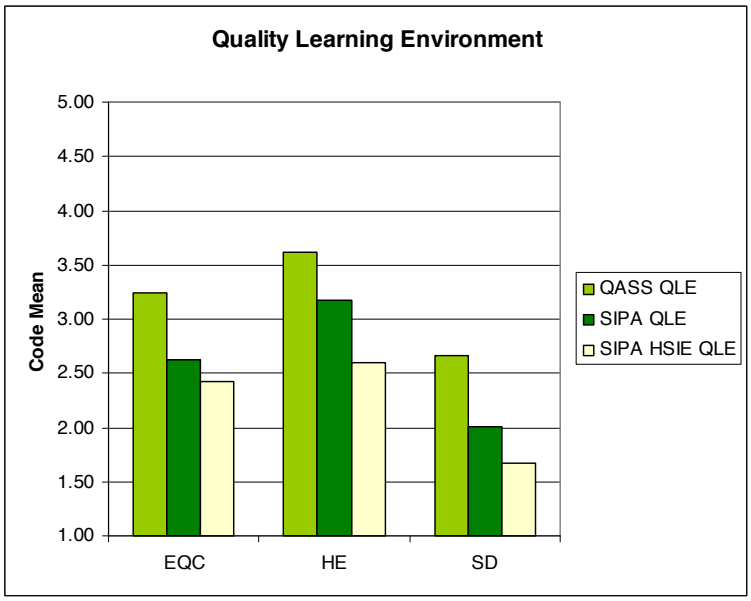

Figure 5. Comparison of task Quality Learning Environment scores among three samples: University social science courses, secondary courses, HSIE courses

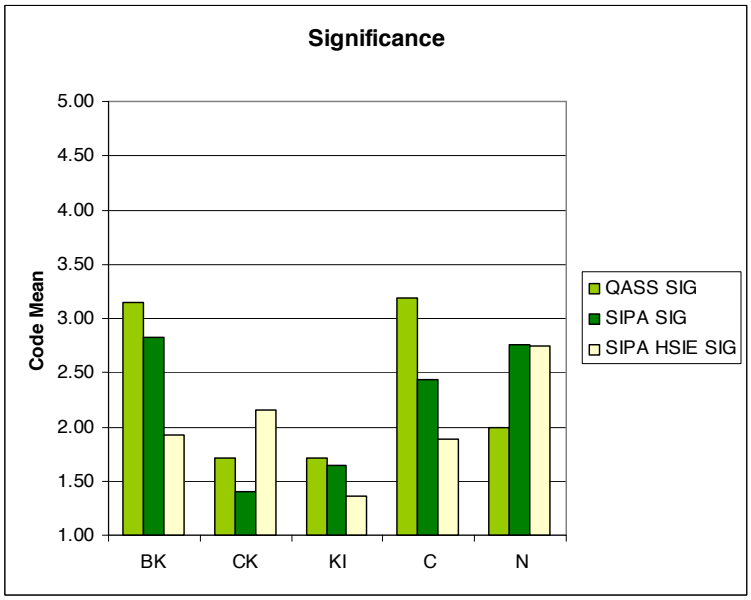

Figure 6. Comparison of task Significance scores among three samples: University social science courses, secondary courses, HSIE courses

Student work quality – 'authentic achievement'

The quality of student work was assessed in two ways for this study. Our primary interest was to assess students' work against criteria known as 'authentic achievement', since these measures have been used productively in earlier studies. Scales for assessing authentic achievement (using a 1-4 rating system) were developed by Newmann and Associates (1996) and modified by Ladwig et al (Ladwig, Gore, Amosa, Griffiths, & Parkes, 2008) for the study of NSW public schools. Studies by Newmann and Associates (1996) have demonstrated that students who perform well on authentic measures also perform well on conventional standardised measures. The quality of student work

collected for this study is summarised in Table 2 and indicates reasonable levels of performance for each element.

QT dimension	Mean (n=233)	SD	Min	Max
Problematic knowledge	1.78	0.78	1	4
Construction of knowledge	2.51	0.94	1	4
Deep understanding	2.67	0.80	1	4
Elaborated communication	2.78	0.79	1	4

Table 2. Student performance using authentic achievement scales

The following examples of student work illustrate a range of codes per element, as measured by the 'Authentic Achievement' scales and demonstrate the differences in student work quality as judged by these measures. The examples have been taken from a variety of courses to indicate how the instrument can apply across social science courses. The selected excerpts from students' work illustrate the kind of judgements made in the coding process but it is important to note that the excerpts were not the sole basis of the given code which was instead dependent on an assessment of the entire piece of student work. In these four examples, the element of authentic achievement that was coded is highlighted together with the course from which the examples come. This information is followed by the question posed for coding purposes, the code given for the student work from which the excerpt has been extracted, and the descriptor for that code. Finally the excerpt from the student work is presented.

Problematic Knowledge: In a Nursing course

To what degree is knowledge presented as problematic?

Coded 3 - Student work treats a moderate amount of knowledge as problematic. Alternative interpretations of problems or events, drawing on multiple sources, are acknowledged and accepted as having equal status.

Student M81Q

Mental health literacy is defined as "the ability to recognise specific disorders, knowing how to seek mental health information, knowledge of risk factors and causes, of self-treatments and of professional help available and attitudes that promote recognition and appropriate help-seeking"(Multicultural Mental Health Australia). A 2001 study in Australia found that whilst 90% of respondents believed mental health is a significant issue, they displayed little understanding of the actual disorders (Francis, Pirkis, Dunt, Blood, & Davis, 2002). It is then logical to assume that improving mental health literacy in the general population would be a positive step toward more tolerant and understanding society. A greater awareness of mental health should, in itself, enable us to a better recognise mental health issues, both in ourselves and in others. A society which is better educated on mental health issues should lead to both earlier interventions and treatments and a greater acceptance of mental illness.

Coded 1 - Student performance treats no knowledge as problematic. All knowledge is presented in an uncritical fashion as not open to interpretation.

Student M81N

The results proved that the participants that completed the course had a better understanding and recognition of mental illnesses, a positive attitude towards people

with a mental illness, agreed with health experts about the treatments, and was more willing to help someone with a mental illness (Jorm & Kitchener, 2002, 2006).
The National Survey of Mental Health and Wellbeing, a community survey completed in 1992 by over ten thousand people, discovered that one in five people have or will have a mental illness (Jorm & Kitchener, 2002). Even though mental illness is common, the general population's understandings about mental illness are incorrect (Jorm, Korten, Jacomb, Christensen, Rodgers & Pollitt, 1997).
It is imperative for the general population to gain an understanding about mental illness, in order to recognise, provide support and reassurance, recommend and encourage people with a mental illness to seek expert assistance, bring mental illness into the open and treat it like other physical conditions, and to bring a end to stigma and discrimination (Jorm, et al, 2006; Jorm & Kelly, 2007; Jorm & Kitchener, 2002).

Construction of Knowledge: In a speech pathology course

To what extent do student performances demonstrate an ability to organise, reorganise, apply, analyse, synthesise or evaluate knowledge, information or text?

Coded 4 - Substantial evidence of construction of knowledge. Most of the student's work includes analysis. At least three statements indicate that the student has correctly generalised, interpreted, tested, or synthesised specific information.

Student F81Q

Roxy's use of language to relay her conversational goal (Bernstein & Tiegerman, 1993) appears incoherent in her language sample. She is unable to communicate what she is referring to at stages e.g. "C: What XX can have?; E: What do you mean?" (#25-26), and is therefore not fulfilling the purpose of her conversation. She also has trouble making some of her conversation topics logical e.g. "C: And hands; A little bit" (#53-55). Her narrative skills are scarce: as she tells the story of the Hungry Caterpillar, she jumps from idea to idea and does not form a coherent story "C: Salami: I don't like salami; I like that; One cupcake." (#95-97). Roxy has a lot of trouble with her use of language both in comprehension and expression; this follows the definition of a disorder. It can therefore be said that she has a pragmatic language disorder.

Coded 1 - Some evidence of Construction of knowledge. A small but not central, portion of the student's work includes analysis. At least one statement shows that the student has correctly generalised, interpreted, tested, or synthesised specific information.

Student F81G

Roxy appears to have disordered syntactic skills based on the transcript provided as this shows that she is inconsistent with her word order. Typically in the English language, word order should be- Subject, verb, object (SVO). However, Roxy confuses this form of language showing use of object, subject, verb (OSV). "Orange this can be".
Based on this information I would diagnose Roxy with a mild language disorder.

Deep Understanding: In a History course

To what extent do students demonstrate a deep understanding of important social science concepts and ideas?

Coded 3 - The student has included social science disciplinary concepts to organise, explain, interpret, summarise or extend the meaning and significance of otherwise discrete pieces of information. The use of the ideas is somewhat limited or shows some flaws in understanding.

Student N81M

The varying opinions in the academic sources examined provide an insight into the nature of history as a constantly changing and often subjective perspective of the past. Whereas "the past" may be defined as the unchangeable events that have

occurred, "history" itself is clearly the subjective perspective that a society and historians place on these events. The version of Australian history omitting Aboriginal history, as presented in Australian history texts between federation and the 1970's, is, therefore, merely one version of history as the society at that time chose to view it. It is, of course, society's perspective of history that is constantly changing- not the facts themselves. C. Healy's suggestion that "Aboriginal narratives are a contribution to historical understanding because they explore the fissures and absences in the European systems of history" will, hopefully, become a more accepted notion by Australian society and enable us to obtain a more accurate version of "Australian History".

Coded 1 - The work includes virtually no social science disciplinary concepts, or the use of any that are included shows almost no understanding.

Student N81l

On some accounts I believe that due to the fact Aboriginal history is oral, unlike written British history it has not been documented in the same way, but there is no denying that we have ignored their existence in this country for many years. This leads me to many questions I would like to ask but am unsure who could provide answers and of fear of being labelled racist, which is not the case at all, I'm not willing to ask. Personally I am proud of my heritage and know where my ancestors are from, though they are not all innocent. If we cannot label ourselves Australians, who are we? I do not condone the mistreatment of Aboriginal people, and this has been further emphasised by these three readings. Australia is my home yet it is also the Aboriginal people's home and a harmonious medium needs to be found. I don't know when this will be achieved or how long it will take or when the Aboriginal people of Australia will be given justice for their suffering.

Elaborated Communication: In an Education course

To what extent does the student performance demonstrate an elaborated account that is clear, coherent and provides richness in details, qualifications and argument?

Coded 4 - The student provides substantial and accurate elaboration for two or more important statements. The details, qualifications, and nuances are expressed within an overall coherent framework intended for the reader, and relevant to the topic.

Student E82J

Expanding on an ethical examination of my teaching, I investigated my lesson for value towards less advantaged students. I had failed in my lesson planning to prepare for students with a disability. Ladwig and King (2003) cite evidence that suggests students with disabilities are best catered for with challenging expectations and high levels of intellectual quality, often outperforming non-disabled students who receive low levels of intellectual quality (King, Schroeder & Chawszczewski, 2001; King, Schroeder & Buckley, in press as cited by Ladwig and King, 2003). Students with low prior achievement also benefit from tasks high in intellectual quality (Newman, Bryk & Nagaoka, 2001 as cited by Ladwig & King, 2003). Furthermore, the NELS analyses (Lee, Smith & Croninger, 1995; 1997 as cited by Ladwig & King, 2003) suggests that lessons high in the Quality Teaching model's Intellectual quality dimension, along with the element of Connectedness (Authentic Pedagogy, Ladwig & King, 2003; NSW DET, 2003), have the power to reduce achievement inequalities between high and low socioeconomic status students. Therefore, the best practice for disadvantaged students is to include them in all activities and ensure high expectations and challenging work designed to achieve high levels of Intellectual quality.

Coded 1 -The student provides virtually no information or provides only disjointed details. Or the student provides discrete claims, broad generalisations, slogans, or conclusions, but none are elaborated.

Student E82A

Sheer content and passion for the content does not necessary lead to a good lesson. Good Inclusivity doesn't necessarily lead to good engagement within the classroom. As the weeks passed the class as a whole was noticeably improving its marking as the markers got closer and closer to marking the same elements with the same score. As the course progressed my understanding of being a teacher was changed. Before this class I may well have been happy in employing the pedagogy of poverty in my classroom.

Course achievement and 'authentic achievement'

In addition to the coding of students' work for levels of authentic achievement, we also accessed the marks awarded for student work by their lecturers. The criteria used in these latter assessments of quality were broad in scope and, as noted above, variable in explicitness. In the following analyses (Table 3 and Figure 7), we depict the relationships between student work quality as assessed within this study (using authentic achievement) and student work quality as assessed within the course (grades/marks assigned by lecturers).

		Authentic achievement	Course achievement
Authentic achievement	Pearson Correlation	1.000	.270**
	Sig. (2-tailed)		.000
	N	233	222
Course achievement	Pearson Correlation	.270**	1.000
	Sig. (2-tailed)	.000	
	N	222	548

**. Correlation is significant at the 0.01 level (2-tailed).

Table 3. Correlation between course achievement and authentic achievement

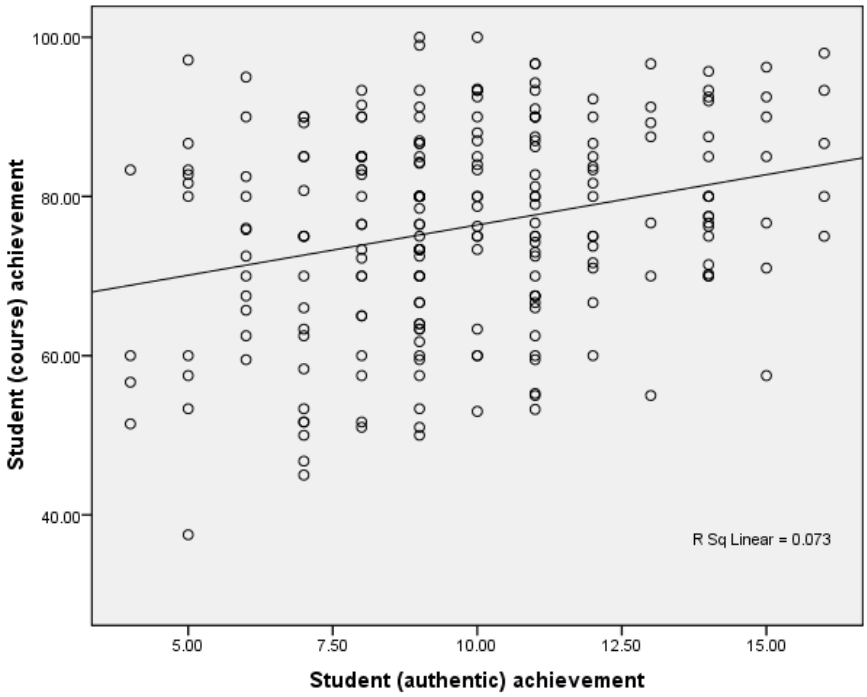

Figure 7. Course achievement and authentic achievement

Figure 7 demonstrates a positive correlation ($r=0.270$, $p<0.01$, $n=222$) between authentic achievement and course achievement. The scatterplot also demonstrates instances where work that rated highly in terms of course achievement sometimes rated very low in terms of authentic achievement. The reverse is not true; that is, higher authentic achievement was not associated with low course achievement. The convergence of achievement at the high end of both 'course' and 'authentic' achievement suggests some consistency between the authentic achievement scales and the expectations of student performance at university level. Further analyses to explore this relationship will include analyses of the contributions of the various dimensions of the QT model to student performance and detailed qualitative analyses of the students' work in order to explore underlying features of the pattern.

Student work quality and task quality

In our earlier research with schools, a strong positive correlation between the quality of assessment tasks and the quality of student work (in terms of authentic achievement) was identified. Figure 8 shows that, at this point in the study, there is no significant relationship between course achievement (grades assigned by lecturers) and overall task quality as measured by the QT instrument. Individual dimensions are significant ($p<0.05$) however, but in the opposite direction to that expected, with Quality Learning Environment and Significance yielding small negative correlations ($r=-0.127$ and $r=-0.152$ respectively) (see Table 4).

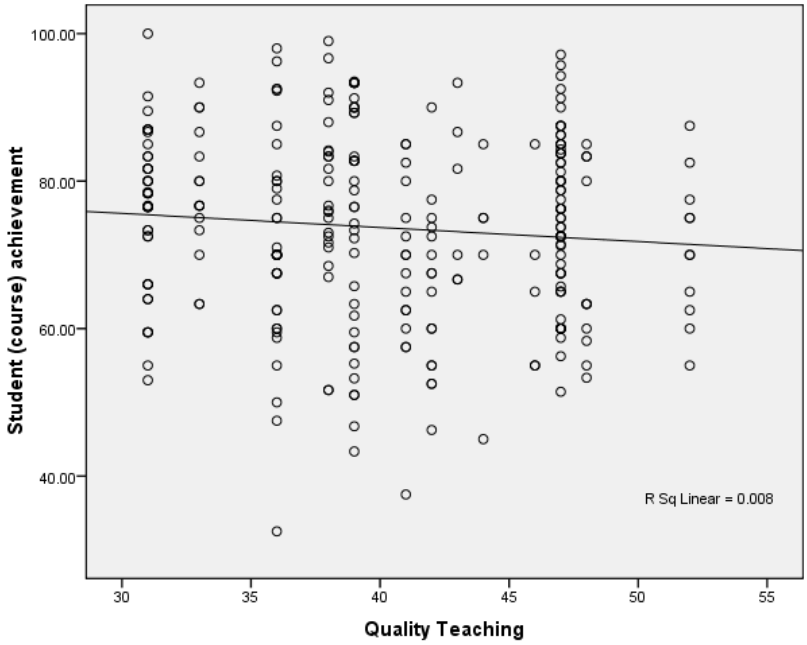

Figure 8. Course achievement and task quality

Furthermore, Figure 9 demonstrates no significant relationship between the 'authentic' achievement scales and task quality at the total QT measure (see Table 4). Individual dimensions are significant with IQ positively correlated ($r=0.182$, $p<0.05$, $n=160$) and QLE negatively correlated ($r=-0.309$, $p<0.01$, $n=160$).

Table 2 Correlations between course achievement, authentic achievement and task quality at the QT dimension level

		Authentic achievement	Course achievement	Intellectual Quality	Quality Learning Environment	Significance	Task Quality Total
Authentic achievement	Pearson Correlation	1.000	.270**	.182*	-.309**	-.099	-.002
	Sig. (2-tailed)		.000	.021	.000	.214	.978
	N	233	222	160	160	160	160
Course achievement	Pearson Correlation	.270**	1.000	.000	-.127*	-.152*	-.092
	Sig. (2-tailed)	.000		.996	.033	.011	.125
	N	222	548	281	281	281	281
Intellectual Quality	Pearson Correlation	.182*	.000	1.000	.199**	.645**	.917**
	Sig. (2-tailed)	.021	.996		.001	.000	.000
	N	160	281	287	287	287	287
Quality Learning Environment	Pearson Correlation	-.309**	-.127*	.199**	1.000	.177**	.411**
	Sig. (2-tailed)	.000	.033	.001		.003	.000
	N	160	281	287	287	287	287
Significance	Pearson Correlation	-.099	-.152*	.645**	.177**	1.000	.849**
	Sig. (2-tailed)	.214	.011	.000	.003		.000
	N	160	281	287	287	287	287
Task Quality Total	Pearson Correlation	-.002	-.092	.917**	.411**	.849**	1.000
	Sig. (2-tailed)	.978	.125	.000	.000	.000	
	N	160	281	287	287	287	287

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

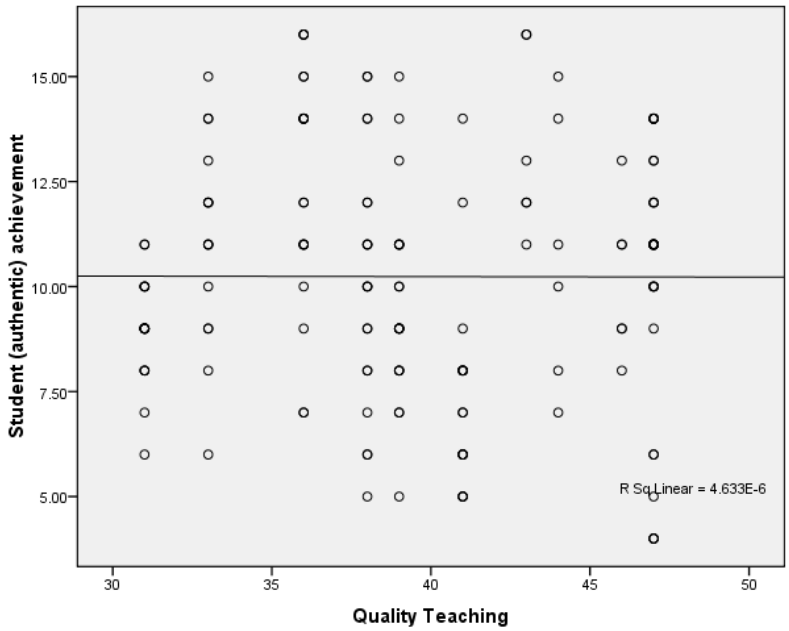

Figure 9. Authentic achievement and task quality

It was anticipated that course achievement would be based on explicit or tacit criteria that range enormously in their alignment with the authentic achievement criteria. We therefore did not necessarily expect a strong correlation between our assessment of tasks and the grades students were assigned for their coursework. However, our work with similar data sets in schools shows a strong positive correlation between task quality (as measured by QT) and student achievement (as measured according to the authentic achievement scales). On this basis a positive correlation between task quality and authentic achievement was anticipated. The 'flat line' result depicted in Figure 9 indicates that even when students are assigned tasks that are low in their intellectual demands, university students are able to produce very high quality work. This finding is not all that surprising given that university students are by definition successful students who have negotiated various forms of assessment successfully and learned to perform to at least a reasonable level on academic tasks.

Conclusion

The types of analysis undertaken in this very preliminary paper will be repeated with the full data set as it becomes available to further explicate the applicability or otherwise of the Quality Teaching model to the rating and refinement of university assessment tasks. Some preliminary findings are as follows:

- the quality of assessment tasks in these university courses is higher than has been typical in secondary schools
- lower ratings for some elements of the Quality Teaching model, such as Problematic Knowledge, than might be anticipated in tertiary education contexts,
- a weak correlation between the marks assigned by lecturers and measures of authentic achievement
- no correlation between task quality and student work in terms of marks assigned by lecturers
- no correlation between task quality and student work in terms of authentic achievement,

As higher quality tasks are produced in the process of refinement that is built into this study, we anticipate that some of these relationships will shift.

References

Amosa, W., Ladwig, J. G., Griffiths, T., & Gore, J. M. (2007). *Equity effects of Quality Teaching: Closing the gap*. Paper presented at the Australian Association for Research in Education Annual Conference.

Boud, D., & Falchikov, N. (2006). Aligning Assessment with Long-Term Learning. Special Issue: Learning-Oriented Assessment: Principles and Practice. *Assessment & Evaluation in Higher Education, 31*(4), 399-413.

Bourdieu, P., & Passeron, J.-C. (1994). Introduction: Language and relationship to language in the teaching situation. In P. Bourdieu, J.-C. Passeron & M. d. S. Martin (Eds.), *Academic Discourse: Linguistic Misunderstanding and Professorial Power* (pp. 1-34). London: Polity Press.

Knight, P. T. (2002). Summative Assessment in Higher Education: practices in disarray. *Studies in Higher Education, 27*(3), 275-286.

Knowles, M. S., Holton, E. F., & Swanson, R. S. (1998). *The Adult Learner: The Definitive Classic in Adult Education and Human Resource Development* (5th ed.). Woburn: Butterworth-Heinemann.

Ladwig, J. G., Gore, J. M., Amosa, W. A., Griffiths, T. G., & Parkes, R. J. (2008). Student performance scoring manual for the social sciences. Newcastle: Quality Assessment: Linking assessment tasks and teaching outcomes in the social sciences.

Ladwig, J. G., Smith, M., Gore, J. M., Amosa, W., & Griffiths, T. (2007). *Quality of pedagogy and student achievement: Multi-level replication of authentic pedagogy*. Paper presented at the Australian Association for Research in Education Annual Conference.

Newmann, F.M. and Associates. (1996). *Authentic Achievement*. San Francisco: Jossey Bass.

NSW Department of Education and Training. (2004). *Quality teaching in NSW public schools: An assessment practice guide*. Ryde: NSW Department of Education and Training, Professional Support and Curriculum Directorate.

Shavelson, R. J. (2007). Assessing Student Learning Responsibly: From History to an Audacious Proposal. *Change, 39*(1), 26-33.