

Refining assessment practice in the social sciences

Jennifer Gore

School of Education, The University of Newcastle, jenny.gore@newcastle.edu.au

Wendy Amosa

School of Education, The University of Newcastle, wendy.amosa@newcastle.edu.au

Tom Griffiths

School of Education, The University of Newcastle, tom.griffiths@newcastle.edu.au

Robert Parkes

School of Education, The University of Newcastle, robert.parkes@newcastle.edu.au

Hywel Ellis

School of Education, The University of Newcastle, hywell.ellis@newcastle.edu.au

Emerging from work on quality assessment in NSW public schools, and drawing upon work in authentic pedagogy and assessment, this paper presents the final analysis from a two-year project, funded by the Australian Learning and Teaching Council, 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.

In this paper we present the findings from our audit of the quality of assessment practice in the social sciences and of the relationships between the quality of assessment tasks and student achievement according to the criteria for performance (a) provided in course materials; and (b) in our authentic achievement scales. Our analyses identified a significant correlation between the quality of assessment tasks and student authentic achievement. As a result of our analyses, we present a case for the use of an effective tool with which to analyse and discuss the quality of assessment practice in the social sciences in the tertiary setting.

Keywords: assessment, authentic achievement, social sciences, quality teaching

Evidenced by the increase in publications and funding opportunities in recent years, assessment in higher education contexts is rapidly being recognised as an area of tertiary teaching and learning that needs dire attention. As argued by Boud (1990) in an early paper on assessment in tertiary education, assessment in higher education contexts has been viewed and treated as necessary only for students to complete so that they can receive certification, rather than, as this paper will argue, to develop authentic assessment that will ultimately benefit students once they have completed their tertiary studies. What Boud has consistently found is a disjuncture between what academics do in their research in terms of applying critical thinking skills; and what they demand of students when completing assessment tasks. Consequently, there is an identified need to set tasks that meet institutional graduate attributes while making explicit authentic connections to students' applications of their learning after they graduate. Therefore, it is argued here, that learning needs to be prioritised when constructing assessment, particularly ensuring that assessment tasks assess the type of learning that is valued in the discipline and professions more broadly.

While substantial energy has gone into understanding the mechanics of assessment in higher education, inadequate attention has been paid to developing means by which university lecturers can monitor the quality of the assessment tasks they develop, a gap that is being addressed through projects such as Boud's Student assessment for learning in and after courses (due for completion at the end of 2009). This paper reports on a research project, funded by the Australian Learning and Teaching Council, 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. The investigation draws on a model for high quality assessment practice, known as the Quality Teaching model initially developed to enhance practice in school education (NSW Department of Education and Training [NSW DET], 2004). The model, which is comprised of three dimensions – Intellectual Quality, Quality Learning Environment, and Significance (see Table 1) – 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.

Table 1. Dimensions and elements of the Quality Teaching model

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

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 contributes to contemporary debates on practices of quality assessment in tertiary contexts. For the work of other contributors on debates regarding the scholarship of assessment, see Boud (2009); Biggs (1996, 2001); Biggs and Tang (2007); and Ramsden (2003).

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.

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, 2009; Boud & Falchikov, 2006; Knight, 2002; Shavelson, 2007) A further criticism of poor quality assessment practice is that it fails to adequately address threshold concepts relevant to specific disciplines (Meyer & Land, 2006), which, as asserted by Meyer and Land (2003) are those concepts which, once learnt, enable students to experience “...opening up a new and previously inaccessible way of thinking about something. It represents a transformed way of understanding, or interpreting, or viewing something without which the learner cannot progress” (p. 1).

This project differs from existing efforts to articulate standards for university teaching often linked explicitly with graduate attributes, 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 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. The research for this project complements the recent work on assessment of Boud (2009) and others by

offering explicit criteria lecturers can use and adapt to their specific subject area and assessment requirements. However, where this research extends on existing work is the assessment criteria used in this project aligns with quality *teaching* approaches already developed and trialled in various education settings (for a discussion on this, see Griffiths, Gore, & Ladwig, 2006); and provides a scaled criteria to determine assessment task quality, rather than a list of statements or questions to which students need to respond.

A variety of taxonomies and checklists are already used in higher education contexts designed for use by academics to make judgements about the quality and appropriateness of tasks assigned to students. For example, Boud (1995, 2009) has developed a number of checklists for this purpose. In addition, Biggs' Structure of Observed Learning Outcomes (SOLO) taxonomy (Biggs & Tang, 2007), a tool used to assess the quality of assessment task objectives, is often used to align assessment objectives with course learning outcomes. The SOLO taxonomy is currently used frequently in Australian higher education contexts for academic development opportunities, such as workshops on assessment quality facilitated by centres for teaching and learning. The work of Boud (1990, 1995, 2009) has consistently emphasised a need for assessment practices in higher education contexts to explicitly align with those educational ideas and values that academics value in their own work. Such practice not only acculturates students into developing deep understandings of discipline threshold concepts (Meyer & Land, 2006), it focuses on the importance of assessment *for* learning as well as *of* learning, relevant to authentic assessment practices. The tool used to assess quality tasks throughout this project enables feedback to be provided to academics in such a way that evidence, in the form of assessment tasks already developed and implemented, is used as the starting point for quality task development.

In this paper, we report findings from our analyses of two research questions that are directly related to the key aims of this project. First, what did our audit of assessment tasks reveal about the quality of assessment practice in the social sciences? Second, what was the correlation between the quality of tasks and student achievement? Throughout our analyses, we identify certain trends in assessment practice in the social sciences that challenge us to reconceptualise what and how we assess in the tertiary sector, in terms of not only what counts as quality assessment, but what counts as quality student achievement. The following section of our paper overviews the methodology of the project.

Methodology

Academics from five disciplines from one university participated 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 Health; and Nursing and Midwifery (courses include prudence in nursing, and mental health literacy and mental health first aid).

There were two phases of data collection for this project. Phase One involved an audit of assessment tasks used by the participants in their teaching of the represented disciplines. This audit included: analysis of the quality of assessment tasks using our measures; analysis of student achievement completed in response to these tasks using our measures of authentic achievement (described below); and professional learning with the participating academics about our tool for analysing the quality of assessment tasks and its evaluation of the tasks they submitted for the project. In the following year and second phase of the study, a second audit of assessment tasks and corresponding student work was conducted. While most of the assessment tasks and student work analysed in Phase Two were from the same courses researched in Phase One, there were some courses for which data was collected only in Phase Two.

Our tool for analysing the quality of assessment tasks was the NSW Quality Teaching Assessment Practice Guide (NSW DET, 2004) which requires judgement on 14 elements which are coded using a five point scale to provide data on levels of Intellectual Quality, Quality Learning Environment and Significance (see Table 1 for the elements and dimensions of the model). The NSW Quality Teaching model draws on the significant work of Newmann and Associates (1996) on Authentic Pedagogy, as well as other elements of classroom and assessment practice that have been linked through empirical research to improved learning outcomes for students across the spectrum of social backgrounds.

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 later modified for research in NSW public schools by Ladwig, Gore, Amosa and Griffiths (2004). The four items that comprise the authentic achievement scales measure: (i) the extent to which students demonstrate disciplinary depth; (ii) the depth of their analysis; (iii) the richness of their communication; and (iv) the extent to which students' work recognises the problematic nature of knowledge. The studies by Newmann and Associates (1996) demonstrated that students who perform well on authentic measures also perform well on conventional standardised measures. Moreover, these scales for analysing student work would appear to be consistent with the emphasis on intellectual depth one might expect in universities.

In addition to the coding of students' work for levels of authentic achievement, we 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 degrees of explicitness.

Findings

This paper reports on the findings of Phase One and Phase Two of our research, describing (i) the overall quality of tasks; and (ii) the relationship between task quality and student achievement. The coding descriptors from our analysis of assessment tasks and student work are provided to illustrate the implication of awarded codes.

(i) The overall quality of tasks

The quality of assessment tasks were analysed using the Quality Teaching assessment instrument (NSW DET, 2004). Table 2 details the mean score, standard deviation and Quality Teaching code (the mean score rounded to the nearest whole number) of the tasks we analysed in both phases of the study for each element of Quality Teaching. Table 2 also provides the descriptor for each of the reported Quality Teaching codes, in order to provide a more complete picture of the judgements made about task quality.

Table 2. Assessment task quality descriptives

	<i>Mean (n=71)</i>	<i>SD</i>	<i>Mean rounded to nearest code</i>	<i>Descriptor for rounded mean code</i>
Intellectual Quality				
Deep knowledge	3.93	1.005	4	The task required sustained focus on key concepts or ideas but did not require articulation of the relationships between and among concepts.
Deep understanding	3.73	0.999	4	A substantial part of the task required students to provide information, arguments or reasoning that demonstrate deep understanding.
Problematic knowledge	2.75	0.982	3	The task required knowledge to be treated as socially constructed, with multiple perspectives addressed.
Higher-order thinking	3.62	1.005	4	A substantial portion of the task required students to demonstrate higher-order thinking.
Metalanguage	2.03	1.069	2	The task required students to make some reference to language, but not how it works.
Substantive communication	3.99	0.918	4	The task required students to present some sustained clarification of the ideas, concepts or arguments related to the substance of the topic.
Quality Learning Environment				
Explicit quality criteria	3.11	1.049	3	Clear statements were made regarding the quality of work but there was little elaboration of what it means to do well.

High expectations	3.66	0.955	4	The task presented a serious challenge for all students.
Student direction	3.11	0.949	3	Students were able to exercise some control in relation to at least one significant aspect of the task.
Significance				
Background knowledge	3.23	0.814	3	Students' background knowledge is mentioned or elicited and is connected to the substance of the task.
Cultural knowledge	2.10	1.333	2	The task requires students to refer to some cultural knowledge but only in a superficial way.
Knowledge integration	2.56	1.262	3	The task requires students to make at least one meaningful connection between topics or subject areas.
Connectedness	3.49	0.808	3	The task is based around some connection to the world outside the classroom, but the task does not require students to explore implications of these connections which remain largely abstract or hypothetical.
Narrative	2.85	1.627	3	The task employs some narrative to enhance the significance of the task and it is connected to the requirements of the task.

The range of the scores for particular elements of the model with and across dimensions is consistent with ratings of secondary and primary school assessment tasks (Amosa, Williams, Ladwig, Gore, & Griffiths, 2008). The overall ratings for most elements were higher than those we found in secondary school assessment tasks across a range of subject areas including secondary school assessment tasks for the subject Human Society and its Environments (HSIE) which is the school subject most similar to the social sciences focus on our study (Gore, Ladwig Amosa, & Griffiths, 2008).

In general, our analysis identified that the assignments appeared to be intellectually challenging, with four of the six elements of Intellectual Quality scoring higher than the possible median score of three. 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. Of particular interest in our analysis of the dimension of Quality Learning Environment was that students were rarely given any detail about what constituted quality when completing the tasks. Explicit Quality Criteria 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 were given no criteria as to what constituted good work. In terms of our analysis of Significance, tasks requiring students to recognise and value non-dominant cultural knowledge (Cultural Knowledge) or to integrate knowledge across subject areas (Knowledge Integration) were particularly rare, with mean scores of 2.11 and 2.62 respectively. The importance of understanding non-dominant cultural perspectives and experiences, 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 across disciplines might be considered desirable in assisting students to develop a coherent body of conceptual and practical knowledge with which to guide their developing professional repertoires and identities.

Of particular interest in this study is not only the quality of tasks according to our measures of quality, but the extent to which the quality of tasks was correlated with the quality of student achievement.

(ii) The relationship between task quality and student achievement

The first component of our analysis of the relationship between task quality and student achievement was to analyse the quality of work that students produced in response to the analysed tasks. Our first mode of analysis involved coding the student work according to our authentic achievement scales. The results of this analysis are reported in Table 3.

Table 3. Student performance using authentic achievement scales

Authentic Achievement scale	Mean (n=976)	SD	Mean rounded to nearest code	Descriptor for rounded mean code
Problematic knowledge	1.97	0.852	2	Student performance treats minimal amounts of knowledge as problematic. The social, political or cultural influences on some content knowledge is acknowledged, but linked to a given body of facts.
Construction of knowledge	2.68	0.903	3	Moderate evidence of Construction of knowledge. A central portion of the student's work includes analysis. At least two statements indicate that the student has correctly generalised, interpreted, tested, or synthesised specific information.
Deep understanding	2.67	0.748	3	The student has included the Social Sciences 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.
Elaborated communication	2.83	0.777	3	The student provides some elaboration for two or three important statements OR provides substantial elaboration for one important statement. In either case, the details, qualifications, and nuances are expressed within a coherent overall framework intended for the reader, relevant to the topic, and without major inaccuracies.

Given the possible range of scores from one to four, and noting that our sample included achievement at the full range of possible scores for all items, the mean scores indicate a reasonable level of overall performance by students.

Our second form of analysis of student achievement focused on the grades that lecturers awarded their students based on course expectations and testing the correlation of this course achievement with our measures of authentic achievement. Our analyses identified a positive correlation between students' authentic achievement and the results students were awarded by their lecturers ($r=0.202$, $p<0.001$, $n=949$), as illustrated in Figure 1.

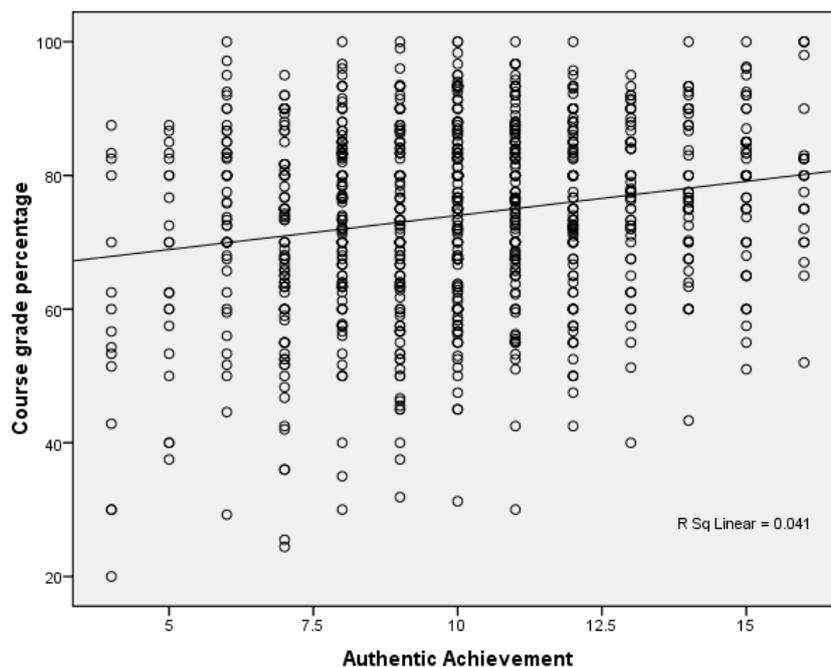


Figure 1. Course achievement and authentic achievement

We 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 the authentic achievement measures and the grades students were assigned for their coursework. The significant correlation found here indicates that, in general, the expectations and judgements of student work made by these university lecturers are aligned with the kinds of standards measured by the authentic achievement scale. Put more simply, on the whole academics tend to agree on what counts as good work, whatever the criteria used in those judgements.

In terms of the relationship between task quality and student authentic achievement, our analyses identified a significant positive correlation between the quality of tasks and the level of student authentic achievement ($r=0.326$, $p<0.001$, $n=946$) as illustrated in Figure 2.

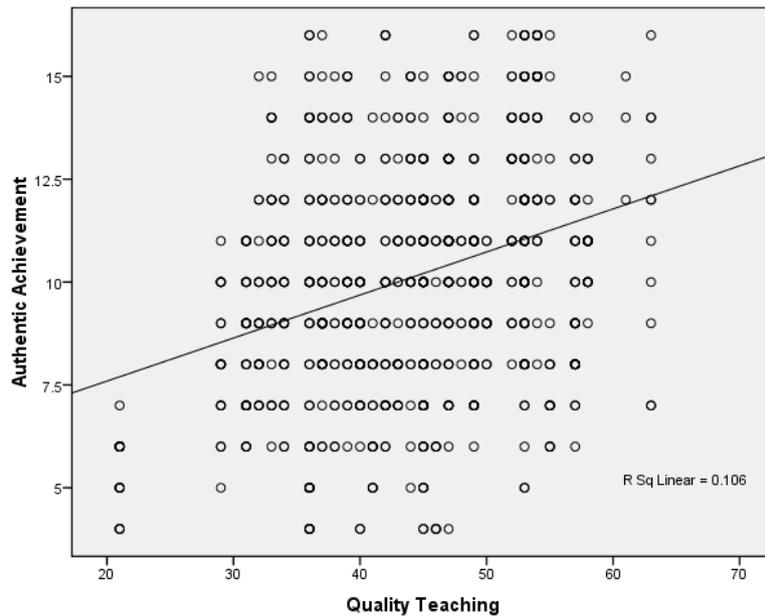


Figure 2. Authentic achievement and task quality

Despite the significant correlation between the quality of tasks and students' authentic achievement, our findings revealed no correlation between the quality of tasks and the results awarded by lecturers ($r=0.005$, $p=.866$, $n=1044$), as illustrated in Figure 3.

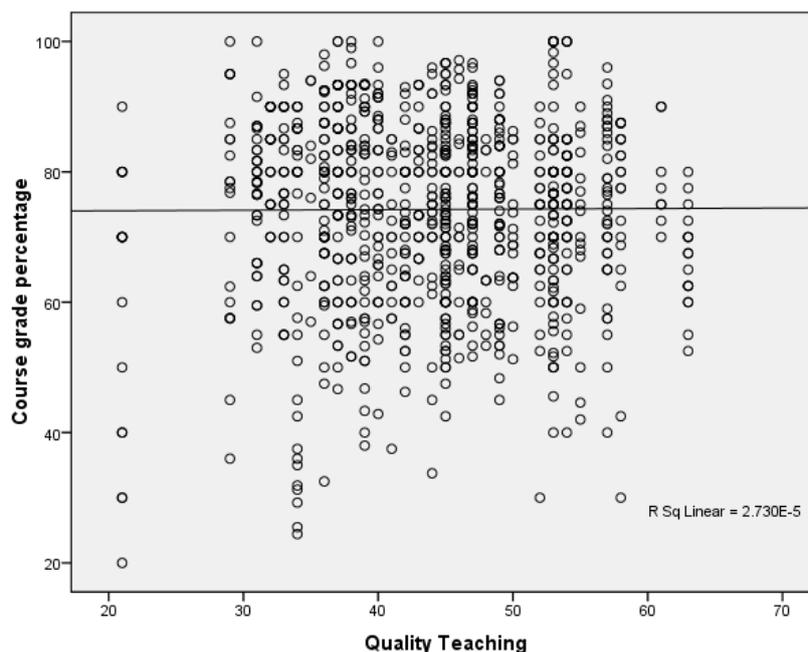


Figure 3. Course achievement and task quality

Our work with similar data sets in schools shows a strong positive correlation between task quality (as measured by QT) and student authentic achievement (Ladwig et al., 2007). The positive correlation between task quality and authentic achievement is evident in this study. In order to more closely examine the relationships between the quality of tasks, students' authentic achievement and their course achievement, Table 4 presents the correlations of students' authentic achievement and course achievement and with the three dimensions that comprise our measure of task quality.

Table 4. Correlations between course achievement, authentic achievement and task quality at the QT dimension level

		Authentic Achievement	Course grade as %	Intellectual Quality	Quality Learning Environment	Significance	Total Quality Teaching
Authentic Achievement	Pearson Correlation	1.000	.202**	.390**	.198**	.154**	.326**
	Sig. (2-tailed)		.000	.000	.000	.000	.000
	N	967	949	964	964	964	964
Course grade as %	Pearson Correlation	.202**	1.000	-.017	.036	.013	.005
	Sig. (2-tailed)	.000		.574	.248	.679	.866
	N	949	1044	1044	1044	1044	1044
Intellectual Quality	Pearson Correlation	.390**	-.017	1.000	.499**	.533**	.889**
	Sig. (2-tailed)	.000	.574		.000	.000	.000
	N	964	1044	1075	1075	1075	1075
Quality Learning Environment	Pearson Correlation	.198**	.036	.499**	1.000	.359**	.668**
	Sig. (2-tailed)	.000	.248	.000		.000	.000
	N	964	1044	1075	1075	1075	1075
Significance	Pearson Correlation	.154**	.013	.533**	.359**	1.000	.819**
	Sig. (2-tailed)	.000	.679	.000	.000		.000
	N	964	1044	1075	1075	1075	1075
Quality Teaching	Pearson Correlation	.326**	.005	.889**	.668**	.819**	1.000
	Sig. (2-tailed)	.000	.866	.000	.000	.000	
	N	964	1044	1075	1075	1075	1075

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

Table 4 reports the significant correlations identified for the Intellectual Quality, Quality Learning Environment, and Significance of tasks with students' authentic achievement, but not with students' course achievement. The lack of any correlation between the quality of tasks, even with any individual dimensions of Quality Teaching, and the students' course achievement is noteworthy. In practical terms, this means that students are sometimes rewarded with high university grades for work they did in response to tasks that asked very little of them – tasks with low intellectual demands, low expectations of quality work, and low levels of significance. It also means that students sometimes receive very low university grades on tasks that have high demands, high expectation and high significance. While the latter finding is consistent with the normal distribution of grades that might be given for challenging work, the fact that the full range of grades is also given for undemanding tasks points to a possible gap in academic skill at assessment task design. At the very least, the result suggests that students might be being assessed on criteria that are not directly related to the intellectual demands of the tasks themselves.

Conclusion

Our analyses identified several key findings that have important implications for further research and for assessment practice in higher education. While the assessment tasks we analysed were of a higher quality than those we have analysed in schools, there were elements, particularly Problematic Knowledge and Explicit Quality Criteria, which scored much lower mean scores than might be considered appropriate in the tertiary setting.

Additionally, while our analysis identified a significant correlation between our measures of authentic achievement and the results awarded to students by lecturers, the exceptions to this trend beg for further analysis into exactly what criteria are being used in the social sciences in universities and the extent to which students are aware of such criteria when completing their responses to assessment tasks. It seems that many students are caught in a game of guess what's in the lecturer's head when it comes to knowing what kind of work will result in good results. While some students are certainly winning this game, our findings suggest that others would certainly benefit from criteria that make explicit what constitutes quality.

In line with our aims of the study, the research reported in this paper found the NSW model of pedagogy to be an effective tool with which to analyse and discuss the quality of assessment practice in the social sciences in the tertiary setting. Our further analyses aim to use participants' feedback of their involvement in the study to examine ways in which the application of a tool such as the Quality Teaching framework might be meaningfully integrated into the professional development of academics in higher education.

In a national context in which there is increasing investment in enhancing the quality of teaching and learning in the higher education sector, we argue that assisting academics in understanding how to develop high quality tasks is a productive direction in which these efforts should move. The Quality Teaching model offers some promise as a tool that can allow for meaningful critique and refinement of assessment development in the social sciences.

References

- Amosa, W., Williams, C., Ladwig, J.G., Gore, J.M., & Griffiths, T. (2008). An Examination of the Quality and Conceptualisation of English Teaching. Paper presented at the *Australian Association for Research in Education Annual Conference*.
- Amosa, W.A., Ladwig, J.G., Griffiths, T.G., & Gore, J.M. (2007). Equity effects of Quality Teaching: Closing the gap [Electronic Version]. In *Proceedings Australian Association for Research in Education Annual Conference*. Retrieved June 23, 2008, from <http://www.aare.edu.au/07pap/amo07284.pdf>.
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32, 347-364.
- Biggs, J. (2001). The reflective institution: Assuring and enhancing the quality of teaching and learning. *Higher Education*, 41, 221-238.
- Biggs, J., & Tang, C. (2007). *Teaching for quality learning at university: What the student does* (3rd ed.). Berkshire: McGraw-Hill Education.
- Boud, D. (1990). Assessment and the promotion of academic values. *Studies in Higher Education*, 15 (1), 101-111.
- Boud, D. (1995). Assessment and learning: Contradictory or complementary. In P. Knight (Ed.), *Assessment for learning in higher education* (pp. 35-48). London: Kogan Page.
- Boud, D. (2009). How can practice reshape assessment? In G. Joughin (Ed.), *Assessment, learning and judgement in Higher Education* (pp. 29-43). Dordrecht: Springer.
- 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. de Saint Martin (Eds.), *Academic Discourse: Linguistic Misunderstanding and Professorial Power* (pp. 1-34). London: Polity Press.

- Griffiths, T.G., Gore, J.M., & Ladwig, J.G. (2006). Teachers' fundamental beliefs, commitment to reform, and the quality of pedagogy. In *Proceedings Australian Association for Research in Education Annual Conference*. Retrieved June 23, 2008, from <http://www.aare.edu.au/06pap/gri06386.pdf>.
- Gore, J.M., Ladwig, J.G., Amosa, W., & Griffiths, T. (2008). On the status and quality of Physical Education as school subject: An empirical study of classroom and assessment practice. Paper presented at the *Australian Association for Research in Education Annual Conference*.
- 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, Australia: Quality Assessment: Linking assessment tasks and teaching outcomes in the social sciences.
- Ladwig, J.G., Smith, M.W., Gore, J.M., Amosa, W.A., & Griffiths, T.G. (2004). *Quality teaching, student performance scoring manual for HSIE*. Newcastle, Australia: University of Newcastle.
- Ladwig, J.G., Smith, M.W., Gore, J.M., Amosa, W.A., & Griffiths, T.G. (2007). Quality of pedagogy and student achievement: Multi-level replication of authentic pedagogy. In *Proceedings Australian Association for Research in Education Annual Conference*. Retrieved June 23, 2008, from <http://www.aare.edu.au/07pap/lad07283.pdf>.
- Meyer, J., & Land, R. (2003). Threshold concepts and troublesome knowledge: Linkages to ways of thinking and practising within the disciplines. Edinburgh: University of Edinburgh. Retrieved August 31, 2009, from <http://www.tla.ed.ac.uk/etl/docs/ETLreport4.pdf>.
- Meyer, J., & Land, R. (2006). *Overcoming Barriers to Student Understanding: Threshold Concepts and Troublesome Knowledge*. London: Routledge.
- Newmann, F.M., & 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.
- Ramsden, P. (2003). *Learning to teach in higher education*. London, UK: RoutledgeFalmer.
- Shavelson, R.J. (2007). Assessing Student Learning Responsibly: From History to an Audacious Proposal. *Change*, 39 (1), 26-33.