

The PEAL project: co-operation or competition in the assessment process

Paper presented at the British Educational Research Association Annual Conference,
Institute of Education, University of London, 5-8 September 2007

David Carless, Associate Professor, Faculty of Education, University of Hong Kong
dcarless@hku.hk

DRAFT: WORK IN PROGRESS

Abstract

This presentation reports on the first year of the PEAL (Primary English Assessment for Learning) project which seeks to enhance the implementation of formative assessment in Hong Kong primary schools. The study uses a case study approach, including interviews with teachers and students, and classroom observations. This paper draws on the notion of the formative use of summative tests, in other words how the testing process can be reconfigured to facilitate productive student learning. Interim findings are discussed in terms of test-related pedagogies, particularly the potential of peer tutoring as a test preparation tool; and tensions in the assessment process: summative versus formative assessment; high stakes versus low stakes assessment; and performance versus learning or mastery. Some implications and issues for further exploration are sketched.

Introduction

This presentation reports on the first year of a 2-year project entitled PEAL (Primary English Assessment for Learning). The project is focused at the interface between assessment and learning, and seeks particularly to highlight the learning function of assessment. Assessment for learning (AfL) is both a global trend sparked by the Black & Wiliam (1998) meta-analysis which claims significant learning gains for formative assessment; and a current initiative in changing the assessment culture in Hong Kong (CDC, 2001). Implementation of AfL in an examination-oriented culture is, however, a challenging proposition with obstacles including teacher attitudes and understandings, and overcoming societal perceptions of assessment as a competitive grading mechanism (Carless, 2005).

The project is addressing two main strands of assessment for learning. Firstly, the formative use of summative tests (Black et al., 2003), in other words how tests can be exploited as a productive learning tool for students. The formative use of summative testing (FUST) seems a contextually suitable strategy in a setting where testing is frequent and pervasive (Carless & Lo, 2006). The second strand relates to peer assessment, self assessment and rubrics (PASAR), in other words the involvement of students in applying standards or criteria to their own work and that of their peers. An important element of PASAR is that such metacognitive processes are also likely to improve student scores in high-stakes tests (Chiu et al., 2007; McDonald & Boud, 2003). PASAR also involves students actively in assessment processes and so carries potential for student empowerment and a positive affective response.

This paper draws on the FUST strand of the project and focuses on the teachers' management of the relationship between teaching and testing. It analyses some of the tensions arising through these processes. Peer tutoring as a learning and test preparation strategy is also discussed. The latter being something that we observed in classrooms and subsequently discussed with teachers, although it was not a strategy that we anticipated seeing or one that is usually associated with regimes of high-stakes testing.

Literature review: assessment and its impact on student learning and affect

This section is divided into four sub-sections which foreground the concepts discussed in the paper. First the relationship between summative and formative assessment is discussed, in particular with respect to the strategy of FUST. Secondly, relevant literature on the impact of high-stakes testing on student motivation is summarized. Thirdly, I discuss the distinction between performance and learning, with the former tending to be more closely associated with summative assessment and the latter with formative assessment. Finally, I briefly review the notion of peer tutoring which data collection revealed as a test preparation activity.

Summative assessment and FUST

Summative and formative assessment are usually distinguished in terms of function and purpose. The former has a primary function of grading or measuring, the latter is about aiding student learning. Newton (2007) queries the traditional formative and summative distinctions and argues that it is more accurate to talk of summative *judgements* (with a variety of purposes) and formative *purposes*. Of course, the same assessment can be used both summatively and formatively, if it provides both an overall grade and learning advice that has potential to be acted upon. In-school tests or examinations often function in this way and the extent to which they include formative dimensions varies from school to school. Such tests might be called micro-summative assessments (William & Thompson, 2006) with the proviso that they will not act formatively unless information that the tests yield is used to modify instruction or student learning.

Given that summative assessment is a necessary reality and that formative assessment is a powerful way of enhancing student learning, are there also arguments for productive synergies between the two? This might accord well with teachers' realities that they need to use assessment both summatively and formatively. Atkin et al. (2005) found that teachers in their project did not make the kind of distinctions between formative and summative assessment found in the literature; grades were a powerful fact of their lives and they could not dissociate them from other elements of assessment. If formative assessment is implemented effectively, it is likely to lead to enhanced performance in summative tests. When summative and formative purposes of assessment are conflated however, there is a danger of summative assessment dominating and formative assessment being neglected (Harlen, 2005).

Students also need to make use of information derived from both summative judgements and formative feedback. Brookhart (2001), for example, reports that good students try to use all assessments formatively and often draw on both summative and

formative data in making adjustments to their work. Student revision is an important part of learning as well as being a test preparation strategy (Dann, 2002).

The ‘formative use of summative tests’, as described by Black *et al.* (2003), provides a well-known example of integrating these two functions of assessment. This notion was stimulated “because their [the teachers’] reality was that formative assessment had to work alongside summative assessment” (Black *et al.* 2003, p. 53). The teachers used methods such as rectifying the learning on test questions done poorly by students; peer marking of test papers; and re-working of examination answers. Similar strategies are also reported in Carter (1997) who found that by shifting responsibility for test performance to learners, they achieved a higher awareness of careless errors and these declined on subsequent tests. Students seemed to demonstrate increased retention of knowledge as evidenced by performance on semester and final exams, and develop better attitudes about test taking and learning.

The impact of high-stakes tests

Summative assessment is often in the form of high-stakes public examinations, i.e. those tests which affect the life chances of students in terms of progression, graduation or the kind of school to which they can proceed. The power of tests is built on perception and so if an exam is believed to be high-stakes, it becomes so (Gipps, 1994). Stakes may also be high for schools and teachers. Barksdale-Ladd & Thomas (2000) report teachers in the U.S. fearing salary cuts or redundancy as a result of low student test scores. Wiliam (2004) generalizes that testing for accountability in the U.S. results in assessments being high-stakes for schools, but low-stakes for students. Heubert and Hauser (1999) warn however that any test that demonstrates to individual students that they are failing is high-stakes because a label of ‘failing’ or ‘low ability’ has deep consequences for self-efficacy and self-esteem. Tests risk sorting students into winners and losers, with the latter entering a cycle of failure and demotivation (Stiggins, 2007). This scenario also occurs in the UK in terms of the pressure of National Curriculum testing (e.g. James, 2000). For example, Pollard et al. (2000) found that even when teachers intended feedback to be formative, students often interpreted it as a summative judgement. There are also dangers of tests demoralising lower achieving students (Harlen & Deakin Crick, 2003). Students with learning difficulties may be deterred by tests and performance targets (GTC, 2004); these lower achieving students often show the greatest gains from formative assessment processes cognitively (e.g. Black & Wiliam, 1998) or affectively (e.g. Miller & Lavin, 2007), but may respond less well to summative assessment.

High stakes summative assessments are necessary for certification and selection purposes but risk narrowing the curriculum through ‘teaching to the test’. Valuable content, skills or dispositions not covered by a test may be neglected. Teaching to the test may also encourage a transmissive mode of teaching that may not suit all students or all situations. Amrein & Berliner (2003) point out that high-stakes tests prompt teachers to take greater control of the student learning experience and deny them opportunities to direct their own learning. Gordon and Reese (1997) also suggest that teaching to the test may enable students to pass the test without having learnt the concepts on which they are being tested.

Performance versus learning/mastery

Dweck (1989) distinguishes *learning goals* or mastery goals in which students strive to increase their competence, to understand or master something; and *performance goals* in which students strive to document gain favourable judgements, or at least avoid negative judgements of their competence. Social actions, such as sharing with peers, are associated with learning goals rather than performance ones (Wentzel, 1996). Performance goals are also believed to be ego-related and stimulate outperforming peers as a way of aggrandizing one's ability status at the expense of peers (Covington, 2000). When performance goals are salient, children who lack confidence in their ability are especially at risk of exhibiting learned helplessness (Dweck & Leggett, 1988). The dichotomous classification of learning and performance goals has been challenged (e.g. Pintrich, 2000) and there is often interplay between these two kinds of achievement goals. Dweck (1999) acknowledges that performance goals are also a societal necessity, but problems arise when proving ability becomes so important to students that it drives out learning goals.

A particular danger in a when a test-dominated context is an emphasis on performance i.e. appearing to be knowledgeable rather than learning or mastery i.e. actually acquiring more knowledge and skills (Ames, 1992). A performance orientation is also associated with short-term learning strategies, such as memorization (Meece, Blumenfeld & Hoyle, 1988). In a culture of frequent testing, there is an obvious risk that students may commit information to short-term memory then forget it shortly afterwards without having achieved productive learning. In short, whilst formative assessment emphasizes learning or mastery, there is a danger that summative assessments over-emphasize performance to the detriment of learning goals.

Peer tutoring

Peer tutoring refers to the use of same age or older students as instructional assistants. Peer tutoring, if implemented effectively, has potential for gains for both the tutor and the tutee (Topping, 2005). The tutor may gain greater understanding of content through explaining it and may also develop a sense of pride and responsibility (Topping & Ehly, 1998). The tutee gains additional support and opportunities to clarify concepts or knowledge. Johnson and Johnson (1983) compared the effects of individual, competitive and co-operative learning experiences and found that the latter resulted in better relationships, higher self-esteem and greater empathy between pupils.

Winter (1996) suggests that peer tutoring may be a particularly appropriate strategy in Chinese contexts, drawing on notions of collectivism and 'human-heartedness'. He suggests that 'Big brother, Big Sister' schemes are useful but result in peer tutoring becoming a break-time or after school activity. He recommends instead peer tutoring within classes amongst same-age peers during standard lessons, so that peer tutoring come part of the regular repertoire of teachers. In a small-scale case study, Greenfield (2003) found that Hong Kong secondary school students were positively oriented towards co-operative learning. Another aspect relevant to the Hong Kong context where teachers' workloads are extremely heavy is that peer tutoring has been

identified as a particularly cost-effective learning strategy (Levin, Glass & Meister, 1987).

The Hong Kong primary school assessment context

In terms of pedagogy, teaching in Hong Kong is generally traditional, text-book oriented and test-driven (Adamson & Morris, 1998). Assessment usually takes the form of tests and examinations, with an emphasis on grades and competition between students (Lee, 1996; Pong & Chow, 2002). There is fierce competition to gain a place at a prestigious secondary school. Secondary school places are allocated through a somewhat complex compromise, whereby internal school results in year 5 and year 6 are scaled against a pre-secondary 1 (year 7) Attainment Test which rank orders schools. Morris et al. (1999), in a study of assessment and feedback, reported that teachers held a set of interlinked beliefs, which stress the validity of objective, reliable, formal assessments and a strong separation between teaching and assessment. They did not see assessment as something which would involve their professional judgement and had a reluctance to assess through any means which might be regarded as non-objective.

Basic Competency Assessments (BCAs) have recently been introduced for the three main subjects of Mathematics, Chinese and English to assess student attainment of the essential knowledge and skills required by students in relation to the learning targets set out in the curriculum for each key stage (CDC, 2006). Progress towards the basic competencies are monitored through the Territory-wide system assessment (TSA) administered by the government at Primary Three (first conducted in mid-2004) and Primary Six (first conducted mid 2005), comprising paper-and-pen mode with an oral assessment component for the two languages. Its stated purposes are to provide feedback to schools about their standards in the three key subjects so that schools could draw up plans to increase effectiveness in learning and teaching. The territory-wide data also help the Government to review policies and to provide focused support to schools (CDC, 2006).

The stated functions of TSA are to help teachers and parents diagnose students' learning, identify their strengths and areas for improvement so that timely assistance can be provided. CDC (2006) claims that the TSA is low stakes in nature and this is true for students because no individual student grades are made available. For schools and teachers however, it is generally interpreted as high stakes because it indicates how well schools are performing. Even the government documents reinforce this: "Monitoring whether individual schools attain the basic standards in key learning areas is premised on the need for accountability" (Education Commission, 2000, p. 46). In other words, the TSA seeks to incorporate both summative (overall assessment data about school performance) and formative (information which can feed back to teaching) functions. The reality however, is a focus on statistics when results are published, and teachers find it difficult to channel the feedback data into classroom implementation (Wan, 2006). This further exemplifies the scenario outlined earlier of formative assessment being drowned by summative assessment. In particular, as year 6 students experience test overload: high-stakes internal assessments; the TSA and then the pre-secondary 1 attainment test all in the same academic year.

Research method

The research strategy involved multi-site case studies. Currently 16 teachers from 10 schools are directly involved in the project with data collection progressing at different rates in different schools. The selection of cases was based on teacher enthusiasm and their potential to contribute to the research questions. In two parallel introductory seminars, a menu of assessment for learning strategies was presented to teachers and they were asked to indicate which ones they were most interested in implementing. The formative use of summative tests was the strategy in which teachers were most interested. This paper reports on work in progress involving data from 5 teachers in 3 schools.

The research method comprised principally semi-structured interviews and classroom observations. A baseline interview was conducted to establish teachers' initial conceptions of assessment and to review existing assessment practices in the schools and as carried out by individual teachers. Interviews were taped and transcribed. Classroom observations were carried out of selected lessons that particularly pertained to the research objectives. Typically a revision lesson was observed just prior to a test; and a consolidation lesson was observed just after a test. A tailor-made classroom observation schedule was used to record classroom processes. Formal or informal interviews with teachers were carried out before and/or after observations to probe issues arising in the lessons.

Formative assessment requires students to act so as to improve their learning. Their perspectives are important and so are tapped through short focus group interviews, informal interactions and through analysis of their written work in quizzes and tests and their performance during classwork.

Data are analysed using established qualitative data analysis procedures. Data reduction is carried out through focused summaries of findings pertaining to the research questions (Miles & Huberman, 1994). The coding of interview transcripts, classroom observation and other data facilitates the development of categories or themes. By developing 'organised constructions' (Holliday, 2002) from the categories, theoretical insights can emerge. Further mining the data seeks evidence or counter evidence that might support or disconfirm the emerging propositions.

Findings

Interim findings are discussed in relation to two themes: peer tutoring as a preparation for in-school tests; and tensions in the assessment process as illustrated by performance versus learning, summative and formative assessment and high or low stakes tests.

Peer tutoring

An example of peer tutoring, which we judged to be successful, occurred in Missionary school, a school which espoused a strong sense of Christian values. Following a workshop on co-operative learning carried out some years previously by an external teacher educator, the Chinese subject teachers started this strategy within

classes using groups of 4 students (one high achieving, two medium and one lower achieving).

A classroom observation of a P4 (year 4) English revision class, prior to a test, indicated Ally, one of the PEAL case study teachers continuing this strategy. Before a group competition when all group members had to go to the blackboard and write answers in response to teacher questions related to test content, the high achieving students were active in coaching their team mates. A later stage of the lesson involved group members helping the lower achieving student in each group in the spelling of the vocabulary items that were being revised. Our observations indicated that students were able to do this quickly, efficiently and with enthusiasm. It did not appear that high achieving students were unwilling to help low achieving ones or that the latter felt labeled or stigmatized. Such strategies might not work with all classes but seem to work well in this case.

Students' views presented a picture of both gains and challenges to working together in groups:

“I like helping each other in the group but I just don't want to be in the group with X”.

“I can help my classmates but it sometimes slows down the progress in the competitions”.

“We sometimes have arguments in our group...then we'd just keep quiet and do nothing ... But we'll work together again after we calm down”.

One of the lower achieving boys who is seated next to and supported by a mature and helpful female student wrote in a Chinese essay about friendship, “A friend is someone who helps me to learn”. The teachers believe that he has derived this notion from the peer tutoring in the class.

Ally commented as follows:

“They really appreciate their peers, so they [lower achieving students] will not feel ashamed. Their acceptance is high ... if their self-image is not good, they may feel embarrassed, but these students accept what they are [i.e accept their abilities/progress]”.

She also reported other peer tutoring strategies: during class, she would ask a student who got the correct answer in a test to explain the process by which she worked out the answer; after giving back marked test papers, she would get students to develop a perfect correction sheet; and during recess time, selected students are paired so that a high achieving one can help a low achieving one. For all three strategies, varying degrees of success were reported.

We returned to the school the following week to observe the lesson after the test as the teacher sought to consolidate student learning of selected test items not handled successfully. Ally described the consolidation days as follows:

Consolidation days after a test are useful ... it gives a message to students that a test or exam is not the end of your learning and that you have to learn from what you have done wrong.

Similar co-operative strategies were pursued but the response from students was much less enthusiastic and whilst some groups were working purposefully, others did not seem to be involved in learning. Ally commented as follows:

Consolidation days are more challenging for the teacher than revision because student problems are different ... After the test they are less motivated because they don't want to look at the test papers again, they think it is ended and now they feel more relaxed

In Catholic primary school, another case study school also with a strong religious ethos, peer tutoring was arranged across rather than within year groups. In this case, year 5 students were particularly deployed to support the lower achieving year 3 and year 4 students. Teri, a case study teacher reported a positive response from both groups of students:

"The year 5 students like to do the peer teaching, some of them even offer to do it every day of the week. The year 3 and 4 students like the peer teaching very much because their big brothers or big sisters really help them and are very kind to them. They will also give them stickers, sweets or potato chips ... their relationships are quite close".

When asked about the extent of learning that was stimulated by the peer tutoring, Teri reported that peer tutoring provided motivation for the students and could be a useful way of doing revision during break time.

"For some lower achieving students, if their parents didn't help them with revision, they might just say 'I don't know anything' and get zero marks for the spelling. But with the help of the big brothers or sisters, they will try to spell the words and at least get one or two correct, and they will feel happy about that".

The tentative indication was that this was a fairly basic form of peer tutoring, mainly involving drill and practice.

Whilst there is still clearly further work that needs to be done, peer tutoring seems to be a strategy worthy of further exploration.

Tensions between testing and learning

A second theme emerging relates to tensions in the assessment process, summative and formative assessment; performance versus learning; and low/high stakes assessment.

Most of the case study teachers espoused views sympathetic to AfL and particularly expressed interest in implementing FUST. They perceived however, a number of barriers as illustrated below:

My colleagues view all testing is summative, they lack the formative concept. ... Teachers have to be aware of using the results and their implications, using the data and not just reporting it.

Because teaching is quite exam-oriented, when students get their results from the assessment, they usually just focussed on one thing: the mark.

I want the assessment to help the students in learning but practically it's quite difficult to do so, because there are too many things we need to do ... the AfL implementation in this school is not very systematic and the linkage between assessment and teaching is not apparent.

Despite the efforts of the government to modify the assessment culture (CDC, 2001; Carless, 2005), the prevailing ethos remains on summative assessment, marks and examination performance, in other words features of a performance orientation.

The theme of performance versus learning also emerged during interviews with teachers. Students were reported as using short-term memory strategies to prepare for tests and even those who achieved reasonably good scores in tests may not have achieved more than surface, short-term learning. Three illustrative quotations from teachers:

“We have noticed that the higher form students [i.e. P5-6] have not adequately consolidated the taught items and they often forget things that have been done before”.

“Many years ago we found dictation is not so meaningful. Some of the students can get very good results in dictation but they cannot use the word”.

“For the lower achieving students, information is stored in the short-term memory before the test, but you really can't see that there is any long-term memory or evidence that they have learnt the material”

There was some evidence that teachers felt under pressure to ensure reasonably good student performance in tests. Two quotations from teachers:

“If the students cannot answer the question, maybe the parents may blame the teachers and query why the students get such a low mark”.

“It is very strange [that] for the challenging questions, maybe the teachers over drill, so even the lower ability students can do most questions. But sometimes for the easy questions, the teachers may neglect it and the students do less well”.

There is a danger that test results are the key performance indicator for teachers and students, but that meaningful learning strategies and interest in learning for its own sake become subjugated.

The notion of performance can be extended from individual students to school performance. In lower primary school, there appear to be no particular significant consequences for students attached to test results, so one might assume that testing was low-stakes. One teacher offered support for this viewpoint as follows: “In primary schools, we have room to do AfL because there are no public exams. Assessment is relatively low-stakes in comparison with secondary schools”. There was also some indication that stakes may be higher for teachers than students. One teacher's views were as follows:

Assessment is for measurement of teachers as well as students to see how well students and teachers have done ... in some ways the tests are low stakes for students, but high stakes for teachers.

Although she acknowledges that the stakes for teachers may only be in terms of perceived pressure or potential loss of face.

Despite the apparent low stakes, there was still evidence that there was a certain amount of pressure on students. Ally reported that “More able students seem to welcome tests and examinations, but the less able seem worried and anxious, probably due to their past experiences of being punished for failing the tests”. Teri reported the difficulties faced by lower achieving students who were reported as finding the tests ‘meaningless’ and often left blank some of the sections without making any attempt. Such students who had accumulated many failures in English were reported as perceiving the subject as “terrible”. Teri described a student as follows: “During the test Stephen just sits there and tells me he doesn’t know how to do it and doesn’t know what the question is about. Honestly speaking he knows what it is asking about but he just wants to sit there, not doing the test”. This echoes the notion of learned helplessness raised in the literature review. In another school, two low achieving students commented as follows: “I’m afraid of the tests and scared of failure” and “I’m afraid of being scolded by the teacher when I get low marks”. Whilst tests may be motivating for some students, for the lower achieving students they may be profoundly demotivating.

The need for good results was particularly evident in terms of the TSA with teachers expending a lot of effort in preparing the students for and practicing the kinds of items required by the TSA. Despite being classified by the government as low stakes, TSA thus has an impact on the design and format of tests taken in primary schools. It is being interpreted as another high stakes external public examinations putting pressure on schools, teachers and students. This is related to the stakes accorded to an examination, such as TSA which may be perceived as even affecting the prospects for survival of the school.

The management of the tension between testing and learning seems to be a particularly urgent challenge. Students may be able to perform quite well in examinations, but may neither remember much of what they have stored in their short-term memory, nor be developing the kind of learning dispositions that may support lifelong learning. Especially at risk are lower achieving students, who may become demoralized and disillusioned well before the end of primary schooling.

Discussion

The data have focused on two themes which have been striking during the first year of the project: peer tutoring as a test preparation strategy and tensions in assessment processes. Peer tutoring as a test preparation activity was an unexpected finding given the generally teacher-centred, textbook-driven nature of pedagogy in Hong Kong. Further ongoing data collection is in process, and the evidence to date modestly enriches the scant existing literature on peer tutoring in Hong Kong. Provisional findings support the notion that peer tutoring may be a suitable strategy in Chinese contexts where collectivism and social relations are particularly emphasised. The strong social bonds between learners seem to be a facilitating factor supporting the

implementation of peer tutoring. Experiences in Ally's class tentatively support Winter's (1996) suggestion that peer tutoring is preferable within class with same age tutors, rather than outside class with different age tutors. Within class peer tutoring can be more frequent and better linked with ongoing classroom learning. The linkage between peer tutoring and preparation and follow-up on tests is not an issue that appears to have attracted much attention in the existing literature, and merits further investigation.

Assessment processes are invariably beset with tensions. A tension prominent in the current discussion is between performance and learning or mastery. The existing literature on performance versus learning mainly focuses on goal-orientations and motivations from a social psychology perspective. Here the emphasis has focused more on the implications for assessment, which have also been discussed frequently e.g. Harlen (2006), Torrance & Pryor (1998). There is a danger that if performance is emphasized throughout schooling then students may accumulate a series of examination results but may not have mastered the examined concepts and may also have little interest in learning for its own sake.

The paper also provides perspectives on the issue of 'stakes' in assessment. Once a 'testing culture' becomes established, there is a tendency to view all tests as high stakes and even an apparently low stakes test in primary school may be seen as an important practice for a future high stakes assessment. When parents, teachers and students are unsure of how test scores may be used or abused, there is a tendency to 'play-safe' and assume that stakes are high. Test anxiety and demoralization of lower achieving students then become significant risks. Within this theme, the study supports similar findings from UK and US, and provides additional evidence of the negative impacts of assessment on lower achieving students.

Conclusions and Implications

The PEAL project is at its mid-point and seeks to develop contextually appropriate formative assessment strategies for the Hong Kong context. Some promising signs are emerging of teacher development in relation to assessment and of good classroom practices, but there is clearly much more to be done in more concretely exploiting the learning potential of assessment. This brief conclusion sets out some key dimensions and outlines some of the issues currently being explored in the second year of the project.

Black and Wiliam (2005) suggest that the effective integration of formative and summative functions of assessment will need to take different forms in different countries, and is likely to be extremely difficult. There is clearly a need for contextually appropriate assessment strategies. In a test-driven setting, such as Hong Kong, there may be a need to link formative assessment to preparation or follow-up on tests. The reality for teachers is that they need to involve themselves with both summative and formative functions of assessment, they are not distinct but overlapping; sometimes competing and sometimes mutually reinforcing. Summative assessment has a powerful impact on the affective response of students, so enabling tests to play a more positive role in the development of productive student learning is also paramount.

The paper has also touched on two themes that appear potentially contradictory: peer tutoring and a competitive performance-oriented system. Peer tutoring is a co-operative strategy popular in Hong Kong schools, and used both to support test preparation and as a learning tool. The data reported here are tentatively indicative of positive affective responses of students towards peer tutoring, without providing any evidence of impact on student learning. It is also suggested that peer tutoring may be a contextually appropriate strategy in Chinese contexts. There is a need to find out more from the case study schools about the processes of peer tutoring and its outcomes.

This co-operative strategy contrasts with a generally competitive performance-oriented examination culture. Issues for follow-up include: to what extent are students in competition with each other or are they allies in competing against an examination-oriented system? How can test preparation become more of a tool for productive student learning as well as a means of accumulating marks; in other words, can learning/mastery be emphasized more than performance? Do performance and learning have to be mutually exclusive? Or to put it another way, if learning proceeds satisfactorily can test performance take care of itself?

A further issue arising from the discussion is a moral imperative to reduce the negative impacts of testing on lower achieving students. How can failing students be supported in the early years of primary education so that they do not fail for the remainder of their school lives? For this group of students, the danger of frequent testing leading to demoralization is particularly strong. Peer tutoring and additional support from teachers may be possible ways forward.

Acknowledgements

I would like to acknowledge my PEAL colleagues Winnie Leung Ka-Wing, Darren Bryant, Jessica Deng Chunrao, Iris Wang Xiang, Daisy Cheng, Sammy Ming, Connie Wong, Lydia Leung, and Professor Liz Hamp-Lyons. I would also like to thank the teachers involved in PEAL activities and the Research Grants Council for their award of funding (HKU 7478/06H).

References

- Adamson, B., & Morris, P. (1998) Primary schooling in Hong Kong, in: J. Moyles & L. Hargreaves (Eds) *The primary curriculum: Learning from international perspectives* (London, Routledge).
- Ames, C. (1992) Classrooms: goals, structures and student motivation. *Journal of Educational Psychology*, 84(3), 261-271.
- Amrein, A. & Berliner, D. (2003) The effects of high-stakes testing on student motivation and learning, *Educational Leadership*, 60(5), 32-38.
- Atkin, J., Coffey, J., Moorthy, S., Sato, M. & Thibeault, M. (2005) *Designing everyday assessment in the science classroom*. New York: Teachers College Press.
- Barksdale-Ladd, M. A. & Thomas, K. F. (2000). What's at stake in high-stakes testing: Teachers and parents speak out. *Journal of Teacher Education*, 51(5), pp.384-397.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Maidenhead: Open University Press.
- Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in education*, 5 (1), 7-74.

- Black, P. & Wiliam, D. (2005) Lessons from around the world: how policies, politics and cultures constrain and afford assessment practices, *The Curriculum Journal*, 16(2), 249-261.
- Brookhart, S. (2001) Successful students' formative and summative uses of assessment information, *Assessment in Education*, 8(2), 153-169.
- Carless, D., (2005). Prospects for the implementation of assessment for learning. *Assessment in Education*, 12(1), 39-54.
- Carless, D. & Lo, Y.C. (2006). The formative use of summative assessment in a Chinese context: synergies or tensions? Paper presented at the APERA conference, HKIEd, November 28.
- Carless, D. (2007). Conceptualizing pre-emptive formative assessment, *Assessment in Education*, 14(2), 171-184.
- Carter, C. (1997). Assessment: shifting the responsibility. *Journal of Secondary Gifted Education*, 9(2), 68-75.
- Chiu, M.M., Chow, W.Y., & C.McBride-Chang (2007). Universals and specifics in learning strategies: explaining adolescent mathematics, science and reading achievements across 34 countries. *Learning and individual differences*.
- Covington, M. (2000) Goal theory, motivation and school achievement: an integrative review. *Annual Review of Psychology*, 51, 171-200.
- Curriculum Development Council (2001) *Learning to Learn: The way forward in curriculum development* (Hong Kong, Government Printer).
- Curriculum Development Council (2006). Basic Competency Assessments. http://cd1.emb.hkedcity.net/cd/eap_web/bca/index3.htm Accessed June 29, 2006.
- Dann, R. (2002) *Promoting assessment as learning: improving the learning process* (London and New York, Routledge Falmer).
- Dweck, C. (1989) Motivation, in: A.Lesgold & R.Glaser (Eds) *Foundations for a psychology of education* (Hillside, NJ, Erlbaum).
- Dweck, C. (1999) *Self-theories: their role in motivation, personality and development*. Philadelphia: Psychology Press.
- Dweck, C. (2002). Messages that motivate. In J.Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education*. San Diego: Elsevier.
- Dweck, C. & Leggett, E. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95, 256-273.
- Education Commission (2000) Learning for life: learning through life. Reform proposals for the Education system in Hong Kong. <http://www.info.gov.hk/archive/consult/2000/Full-Eng.pdf>
- General Teaching Council (2004) Internal and external assessment: what is the right balance for the future? Advice to the secretary of state for education and others. <http://www.gtce.org.uk> Accessed July 24, 2006.
- Gipps, C. (1994) *Beyond testing* (London, Falmer Press).
- Gordon, S. & Reese, M. (1997) High stakes testing: worth the price? *Journal of School Leadership*, 7, 345-368.
- Greenfield, R. (2003) Collaborative e-mail exchange for teaching Secondary ESL: A case study in Hong Kong, *Language, learning and technology*, 7(1), 46-70.
- Harlen, W. (2005) Teachers' summative practices and assessment for learning – tensions and synergies, *The Curriculum Journal*, 16(2), 207-223.
- Harlen, W. (2006) The role of assessment in developing motivation for learning. In J. Gardner (Ed.) *Assessment and learning*. London, Sage
- Harlen, W. & Deakin Crick, R. (2003) Testing and motivation for learning, *Assessment in Education*, 10(2), 169-207.
- Heubert, J. & Hauser, R. (Eds.) (1999) *High stakes: testing for tracking, promotion and graduation*. Washington, DC: National Academic Press.
- Holliday, A. (2002). *Doing and writing qualitative research*. London: Sage.
- James, M. (2000) Measured lives: the rise of assessment as the engine of change in English schools. *The Curriculum Journal*, 11(3), 343-364.

- Kennedy, K. et al. (2005) Assessment for productive learning: forms of assessment and their potential for enhancing learning. Paper presented at the 32nd Annual Conference of the International Association for Educational Assessment, Singapore, 21-26 May.
- Johnson, R. & Johnson, D. (1983) Effects of co-operative, competitive and individualistic learning experiences on social development. *Exceptional children*, 49(4), 323-329.
- Lee, W. O. (1996) The cultural context for Chinese learners: conceptions of learning in the Confucian tradition, in: D. Watkins & J. Biggs (Eds) *Teaching the Chinese Learner: psychological and pedagogical perspectives* (Hong Kong, Comparative Education Research Centre), 25-41.
- Levin, H., Glass, G. & Meister, G. (1987) A cost-effectiveness analysis of computer-assisted instruction, *Evaluation Review*, 11(1), 50-72.
- McDonald, B. & Boud, D. (2003) The impact of self-assessment on achievement: the effects of self-assessment training on performance in external examinations, *Assessment in Education*, 10(2), 209-220.
- Meece, J., Blumenfeld, P. & Hoyle, R. (1988) Students' goal orientations and cognitive engagement in classroom activities. *Journal of educational psychology*, 80, 514-523.
- Miles, M. & Huberman, A. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage.
- Miller, D. & Lavin, F. (2007) 'But now I feel I want to give it a try': formative assessment, self-esteem and a sense of competence. *The Curriculum Journal*, 18(1), 3-25.
- Morris, P., Adamson, B., Chan, K.K., et al. (1999) *The project on feedback and assessment: Final report*. Hong Kong: Hong Kong University.
- Newton, P. (2007) Clarifying the purposes of educational assessment. *Assessment in Education*, 14(2), 149-170.
- Pintrich, P. (2000) Multiple goals, multiple pathways: the role of goal orientation in learning and achievement. *Journal of Educational Psychology*, 92(3), 544-555.
- Pollard, A. & Triggs, P. (2000) *What pupils say: Changing policy and practice in primary education*. London: Continuum.
- Pong, W.Y. and Chow, J.C.S. (2002) On the pedagogy of examinations in Hong Kong, *Teaching and Teacher Education*, 18 (2), 139-149.
- Stiggins, R. (2007). Assessment through the student's eyes. *Educational Leadership*, 64(8), 22-26.
- Topping, K. (2005) Trends in peer learning, *Educational Psychology*, 25,(6), 631-645.
- Topping, K. & Ehly, S. (1998) Introduction to peer-assisted learning. In K. Topping and S. Ehly (Eds.), *Peer-assisted learning*. Mahwah, NJ: Lawrence Erlbaum.
- Torrance, H. & Pryor, J. (1998) *Investigating formative assessment* (Buckingham, Open University Press).
- Wan, W. P. (2006) Washback effects of basic competency assessment in Hong Kong primary English language curriculum. Unpublished M. Ed assignment, University of Hong Kong.
- Wentzel, K. (1996) Social and academic motivation in middle school: concurrent and long-term relations to academic effort. *Journal of Early Adolescence*, 16, 390-406.
- William, D. (2004). *The validity and impact of educational assessments*. Paper presented at the Discussion Group 12 at the 10th International Congress on Mathematical Education. Copenhagen, Denmark.
- William, D. & M.Thompson (2006) Integrating assessment with learning: what will it take to make it work? In C.Dwyer (Ed.), *The future of assessment: shaping teaching and learning*. Mahwah, NJ: Lawrence Erlbaum.
- Winter, S. (1996) Peer tutoring and learning outcomes. In D. Watkins & J. Biggs (Eds.), *The Chinese learner: Cultural, psychological and contextual influences*. Hong Kong: Comparative Education Research Centre and Australian Council for Educational Research.