Submission Number 224: Benefits for students from achieving accuracy in criteria-based self-assessment.

Danny Carroll, University of New South Wales, Australia
d.carroll@unsw.edu.au

Abstract:

The student experience of assessment is too frequently a passive one. At the University of New South Wales (UNSW), Australia, the increasing use of self-assessment through an online criteria based marking and feedback system (Review) is improving student's experience of assessment. This mixed methods study examines student's attitudes to and accuracy in self-assessment against criteria. Select cohorts practiced calibrating self-assessment judgments as a pre-step to doing self-assessment. Students reported positive attitudes to this activity and that practice contributed to forming better professional judgment. Of the three cohorts studied, the class incentivised with bonus marks for practicing self-assessment and accuracy in self-assessment achieved the highest participation rates and accuracy. Through regression analysis of self-assessment accuracy inside tasks, accuracy was found to be significant against course marks. Characteristics of improved self-assessment accuracy by accuracy bands are analysed with evidence that benefits are most marked at the extreme over assessor band. Improvements in self-assessment accuracy were also observed in increased percentages of accurate assessors and moderation among extreme under-assessors. The results overall highlight benefits for students of embedded self-assessment practice and the need to better integrate self-assessment practice into our Program (degree) design.

Keywords: self-assessment accuracy, criteria based marking, developing professional judgement

Assessment drives learning and a concomitant requirement of assessment is feedback. Studies tell us that students value and use feedback; local course evaluations show an almost insatiable desire for ever increasing amounts of feedback. Feedback is most effective when it satisfies the conditions of good practice with cognitive engagement and learner investment in the task and assessment process (Nicol, 2006). Many strategies can be employed to engage learners in assessment, but one of the most immediate, enduring, and personal is the practice by individuals of meaningful self-assessment.

Too often, assessment in Higher Education doesn't generate effective learning. Assessments are experienced as compliance and certification exercises, box-ticking hurdles for progression. Assessments are rarely linked explicitly to the development of student's discipline knowledge, skills and professional predispositions. Individual tasks (and even whole courses) have little sense of articulation into degrees. The student's experience of assessment is too often of an act 'done' to them, rather than one which they participate in and clearly identify they benefit from. Feedback is often wasted, as it's provided too far from the learning moment, is not 'strategic' to the learner's main needs or linked to improving.

21st century learning demands learners to be visibly engaged with their discipline. It is generally no longer enough to be a content master; modern graduates need to be able communicators, team-workers, have international outlooks, be critical thinkers, and possess demonstrable reflective skills (OECD, 2008). Contributing to this, I argue that the endemic practice of self-assessment makes a powerful and still under-used contribution. The encouragement of a pre-disposition towards accuracy in recognising 'quality' (in self and others) is a graduate outcome Higher Education providers should aspire to.

The rationale and design of this research is informed by learning and cognition theories including constructivism, social cognitive theory (Bandura, 1994 & 2001), and professional competency development. Research from the discipline of Psychology related to metacognition, optimism bias and self-protective behaviours also informs this research. In addition, the rich body of work relating to student self-assessment practice was referenced, with David Boud's extensive contributions a constant inspiration. Indeed, Boud (2003) and Rust (2003) have noted the need for more empirically based studies in real-world situations where the benefit to students of self-assessment practice is the focus, a call this research responds to.

Assessment is increasingly required to do 'double-duty', to serve both as a valid and reliable indicator of student achievement and for its consequential validity - as a springboard for learning. Assessment is most beneficial for students when it not only accurately reflects a sound judgment of their performance on a particular task, test or undertaking, but also provides a structure for effective, long-term learning. Yet, assessment designed 'for' learning, is still, arguably, not a mainstream reality (Manitoba Education, 2006).

The desirability of assessment design that assists learners to learn and generates longer-term knowledge, cognitive or behavioural changes has been extensively studied (Nicol, 2007, Shepard, 2000). One beneficial teaching and learning activity is student's practice of self-assessment, which can be operationalised in a range of ways. Self-assessment motivates students to engage with tasks and assists them to focus on the performance standards and issues of quality that expert judges 'know' (and as a result, often judge holistically).

This study focuses on student criteria based self-assessment judgments. Criteria based judgments are particularly valuable because they:

- make explicit to novice learners what expert judges value
- scaffold performance targets to students
- interact with rubrics and exemplars to establish 'quality'
- provide a framework for improving self-regulation and reflection in action

The relative inaccuracy of student's self-assessment has diminished the worth of self-assessment for some, as it is, for a variety of reasons, demonstrably a less valid and reliable measurement tool of performance than the outcomes of standardised tests and expert markers. More interesting however than the validity and reliability of student self-assessment as a hard measure, are the associated questions. Why are students inaccurate in self-assessment, what do different presentations of accuracy and inaccuracy potentially tell us (and the students), how can we support them best to practice and develop accuracy? Above all, what do students 'get' out of practicing self-assessment with an accuracy focus?

Additionally, it should be noted that many of the studies of self-assessment have been conducted at the level of student estimates of course or overall task performance (Ross, 2006). Estimations in these contexts are not necessarily transferrable to research findings on student self-assessment accuracy against criteria judgments. The outcomes are vastly different as well. As outlined above, if it's accepted that criteria scaffold the thought of experts to novices, then learner engagement with criteria supports cognitive engagement processes in a way that task and course final mark prediction activity cannot. They assist learners to adapt the mental models of expert judges in a particular sub-domain area.

There are numerous reasons to expect inaccurate student self-assessment. Firstly, the practice of self-assessment against criteria is still a relatively uncommon component of our teaching

(Carroll 2013). Additionally, while many educators model performance standards and teach towards quality exemplars, just as many do not (Hattie, 2008): therefore students often learn with inadequate models of quality standards to perform towards and judge against. In addition, psychological optimism biases and self-protecting behaviours are complicit in acting against objective self-assessment (Dunning, Heath & Suls, 2006). Even high-performing students (who one might intuitively expect to be relatively accurate) are generally inaccurate (under-assessing) self-assessors. Despite a sound discipline knowledge base they seem to lack the contextualised calibrating experience or knowledge of where particular standard lie in relation to judging their performance.

The differential presentations of student self-assessment inaccuracy are informative as well, both for what they tell us about general populations and for the tantalising individuals that 'buck the trend'. Studies of student's self-assessment accuracy estimates show students accuracy tends to 'band' against dependent variables of task and course marks. Boud (2003) outlines the commonest scenario: high achieving students tend to underestimate their achievement, and a larger cohort of low-performing students tend to strongly over-rate their work. Mid-level achieving students tend to plump for average self-assessment estimates.

While the trends are clear, less clear is the relationship between interventions to encourage accuracy and the size effects, speed and persistence of any benefits that students get from practicing self-assessment. In relation to improved self-assessment accuracy, what are the outcomes for individuals if self-assessment improves? Are these benefits generally uniform in nature, with common characteristics experienced by most students and clearly observable in effects on different 'bands'; or are the outcomes so idiosyncratic that generalizability is impractical?

Boud, Lawson & Thompson (2013) advanced empirical evidence of student's improved self-assessment accuracy through repeated practice in a longitudinal study based on repeated self-assessment practice. The study paints a picture of gradual movement towards accuracy through degree progression, and while agreeing with the conclusions found, it should be noted that the sample of third year practitioners is small compared to the number of observations of first and second year self-assessors.

Intuitively, it's appealing that final year undergraduates should be more accurate at self-assessors than first-years. They have had more opportunities to calibrate their work against the feedback and marks from numerous assessments. They will have formed more evidence–based views of what is quality work in their discipline and the relative worth of their work. As well, they will have views about themselves as practitioners and are more likely to have developed beneficial self-regulatory mechanisms and self-reflective behaviours.

The imminent approach of professional practice, helps focus final year students on the development of self-regulating practices, including self-critiquing and reflection. Reflective practitioners and a predisposition to be 'objective' in self-assessment are highly desirable attributes for graduates.

Context of study

The long-term research plan is a longitudinal study of student accuracy in and attitudes to criteria based self-assessment in Business School cohorts. The study examines if there are measurable and / or self-reported benefits of teacher interventions designed to improve student focus on and accuracy in self-assessment. The study seeks to better understand both the patterns related to self-assessment accuracy and what improvement can mean for individual students development of self-regulation, self-efficacy and cognitive change.

Lecturers in Charge (LICs) of a first year postgraduate, a second year and third year undergraduate course in the Australian School of Business (ASB) at UNSW, Australia, hosted an ethics approved study in 2012-13. Initial findings on student attitudes to self-assessment (which were highly positive) were reported on in a separate paper (Carroll 2013).

Each of the three courses marked two tasks through Review, an online criteria based marking software. Review provides an online interface for markers to mark and comment against criteria and for students to also self-assess against the same criteria. Students could self-assess more than once and were encouraged to use the self-assessment process to check progress to task and criteria fulfilment.

Aim:

The aim of this paper is to observe relationships and covariance between student self-assessment activity (over confidence, accuracy and under confidence) and task and course marks outcomes. Specifically, the statistical analysis considers the significance and effect of student's positive and negative self-assessment estimates. Descriptive statistics are also analysed to identify patterns representing common outcomes for 'types' of student behaviours.

Method
The overall research is a mixed methods design on student attitudes, practice and accuracy in criteria based self-assessment embedded in course tasks. Components of the design have included the student's optional participation in surveys, the option to do embedded self-assessment (in Review) and supportive activities (in-class activities, lecture visits and emails) to assist students practice and develop accuracy self-assessing against criteria.

Factor Analysis of the survey data resulted in consistent, reliable factors demonstrating strong student identification with the practice and utility of practicing self-assessment and developing professional judgments through this process. Additionally, qualitative analysis of student's self-reported experience was generally very favourable (Carroll, 2013).

Relevant to this paper, student's criteria self-assessments, task and course achievement data were compiled in a data set. The set was then anonymized prior to simple regression tests for significance and effects size between dependent (task and course outcomes) and independent variables (student's self-assessment accuracy). Regression tests were carried out against task 1 and task 2 self-assessment accuracy against the dependent variable (the task final mark) in SPSS: results are reported below. In addition, descriptive statistics on the student population and performance 'bands' were conducted in Excel. The data of non-self-assessing students was excluded from further analysis.

Student self-assessment accuracy data is represented in a number of ways. Observations about students are studied in regressions in terms of their absolute distance from accuracy (agreement with tutor's mark). They are also analysed as over (positive / over-optimistic) and under-assessors (negative or lower self-assessment estimates against the mark assigned). Band' systems of accuracy are used and labelled as presented. The bands are based around a shared definition in this study of 'accuracy', where student's estimates fall within + or – 5 of the assigned mark.

Staff and student criteria judgments are placed at any point from 0 to 100 on a visual slider scale in the Review software (see Figure 1). Each point on the slider continuum represents a number mark, which is typically hidden from student, but familiar performance grade boundaries (F, P, C etc) are visible. Criteria can be weighted and Review totals the weighted

criteria mark into a final task mark score. In this study however, unweighted criteria marks by staff and students have been used to ensure an unweighted basis to assist analysis.

Figure 1: Review marking slider

Staff and students are informed that staff can't see student's self-assessment estimates until after marking is completed. All staff marks and student self-assessment predictions are stored in a centralised database and marks and feedback are published online.

Process:

Pre-semester teaching team meetings were held in each course. A brief explanation of the research and its theory base was provided. More time was spent on the rationale behind the feed-forward of information to students and the proposed in-class practice. This involved student discussion and analysis of criteria and quality exemplars to be conducted by tutors. LIC's related this activity to marking standards of course tasks.

The inputs to student to encourage self-assessment activity and accuracy improvement were 'feed-forward' lecture visits, socially based in class discussion activities, supportive reinforcement emails and encouragement from teachers to self-assess. As part of the feed-forward process, students were advised in the lecture and via a summary email to avoid over-optimistic or aspirational self-assessment. To control for the misconception that over-optimistic self-assessment might influence markers to inflate their mark, students were informed that their self-assessment estimates were only visible after marking is completed. Therefore high self-estimates couldn't influence the markers judgment. Some observations about previous similar cohorts over-optimistic self-assessment estimates and task level outcomes were presented to students prior to the first assessment task in an attempt to establish an early reference base inclining the students towards rigour.

The task criteria were similarly formed and consistent from task 1 to task 2. Between task 1 and task 2, students enjoyed the feedback and self-assessment marks from the first task as an aid to calibrating their estimates and judgment for the second task.

Table 1: Task self-assessment, survey completion, demographics

	Course	Students in course	% Self-assessed Task 1	% Self-assessed Task 2	% Did Survey	% Female / Male	% Under 25	% Local	% NESB
A	1st year PG	131	73	60	73	59 / 41	74	25	85
B	3rd year UG	622	42	25	19	61 / 39	98	53	68
C	2nd year UG	178	91	74	85	62 / 38	99	59	50

NESB = Non-English Speaking Background.

Students undertook voluntary self-assessment in two similar criteria based tasks in the three courses below (2 undergraduate and 1 postgraduate). In the two courses where self-assessment activity was incentivised, the rates of student participation increased markedly. In the 1st year postgraduate course (Course A - see Table 1), a 1% bonus mark for completing self-assessment activity and this boosted self-assessment activity markedly above the voluntary (and unrewarded) model of the 3rd year undergraduate course (Course B). In the 2nd year undergraduate course (Course C) a 1% bonus mark for completing self-assessment in each task was offered and 1% bonus mark was awarded to students who were accurate in self-assessment. Incentivisation had a dramatic effect, with the best survey completion, self-assessment participation and self-assessment accuracy rates being achieved in the most incentivised course. This course also had the best overall 'embedding' of self-assessment support and practice, due to the small, enthusiastic teaching team.

In all courses, females outnumbered males, the cohorts were youthful, Chinese nationals represented the largest nationality grouping outside of Australians and among the local students, many reported having Non-English-speaking backgrounds.

Findings:

In the associated survey, students predicted their final course grade. Predictions were over-optimistic, with the first year postgraduate students (many of whom were recent arrivals in Australia) being more over-optimistic than the undergraduate students. Second year undergraduates provided the most realistic estimates (see Table 2). It's notable how few students predicted a Pass grade: unsurprisingly no students in any course predicted a Fail (F).

Table 2: Students predicted and actual course grades

A – 1st year PG (surveyed, n = 93)			B - 3rd year UG (surveyed, n = 119)			C - 2nd year UG (surveyed, n = 151)		
Course Grade	% students - grade predicted	% students - actual grade	% students - grade predicted	% students - actual grade	% students - grade predicted	% students - actual grade		
F	0	1.04	0	5.12	0	0.56		
P	1.1	25.77	3.4	48.20	4	12.29		
C	11.8	48.45	31.1	27.90	47	47.49		
D	57.0	20.62	50.4	10.22	44	39.11		
HD	30.1	4.12	14.3	8.56	5	0.56		
Invalid	0	0	.8	0	0	0		

Grades range from Fail (F) to High Distinction (HD)

Findings:

In regressions, student self-assessment accuracy was represented by the difference between student's predicted criteria mark and the staff (unweighted) criteria mark. The dependent

variable is final task mark. Regression models were used to test the impact of self-assessment accuracy on task performance.

Table 3: Regressions Task 1 & 2

	Course A - 1st Yr	Course B - 3rd Yr	Course C - 2nd Yr
Task 1			
Co-efficient	-0.33***	-0.11**	-0.53***
Standard error	(-0.088)	(-0.051)	(0.054)
Observations	96	246	162
R Squared	0.12 (adjusted R)	0.02	0.37
Task 2			
Co-efficient	-0.48***	0.03 & not sig.	-0.43***
Standard error	(0.120)		(0.065)
Observations	96	156	132
R Squared	0.11 (adjusted R)		0.25

*,**,***denotes significance at 10%, 5% and 1%. The dependant variable is final task mark (unweighted), the independent variable is student self-assessment difference from final mark.

Most task regression results showed significance and strong effect sizes. Students who underestimate their performance (via a negative self-assessment estimate) tend to achieve a better average final task mark than students who over-estimate their self-assessment. In some cases the coefficients showed strong effects (see Table 3). The 2nd year undergraduate (Course C) task 1 and 2 showed a positive or negative movement of an average .53 and .43 of a mark on the final task mark for every 1 mark plus or minus in student's self-assessment accuracy estimate.

In the 1st year postgraduate course (Course A), the regression run on student self-assessment difference against final task mark was significant in both cases. When a regression was run on the 'improvement differential' of improved self-assessment accuracy from task 1 to task 2, this was not significant (the dependant variable in this case was the final overall course mark). However, students who under-assessed themselves overall on both tasks did exhibit significantly higher final course marks than other students. In the combined self-assessment estimates from the two tasks, 17 students out of 96 recorded overall negative or under-assessing estimates compared to the markers. Negative self-assessing students averaged 72.53 as their final course mark while the over-assessors final course mark average was 64.54.

Table 4: Task student self-assessment accuracy 'bands' and marks in course tasks

Course / Task	Obs	+ Over-assessors %				Accurate assessors %				- Under-assessors %			
		>+10	Mark	<10 - >5	Mark	>5 - 0	Mark	> 0 to -5	Mark	> -5 to < -10	Mark	> -10	Mark
A1	107	38.3	62.9	24.3	63.8	**19.6**	68.9	**13.1**	70.4	1.9	74.5	2.8	83.3
A2	96	25.0	62.9	26.0	65.7	**29.2**	71.3	**14.6**	73.5	3.1	74.2	1.0	71.3
B1	260	25.0	65.7	17.3	70.3	**29.2**	74.6	**17.3**	76.0	6.2	79.9	5.0	79.9
B2	157	21.0	65.9	17.2	72.8	**18.5**	76.0	**15.3**	76.6	15.9	79.8	12.1	83.6
C1	163	10.6	61.4	19.3	68.3	**34.8**	71.0	**19.9**	70.6	9.9	73.0	6.2	74.6
C2	133	6.1	63.9	15.2	74.4	**33.3**	74.1	**28.8**	76.7	10.6	74.6	6.0	79.8

The task mark is the average task mark for all the students in the self-assessment band in each case above.

More uniform tutorial instruction on analysis of exemplars of quality was received in courses A and C. In Course B, the large number of tutors made a consistent class based experience of the self-assessment supporting activity hard to achieve. This is likely to have affected the rigour and accuracy of the practice of self-assessment in these courses. Certainly, Course A and C evidenced larger effect sizes in regression and higher and more consistent participation rates in self-assessment activity by students. These were the courses where self-assessment activity was rewarded with bonus marks.

It was notable that the best participation rates and self-assessment accuracy occurred in the 2nd year course (C), where there was comprehensive implementation of student analysis and discussion of exemplars of quality work. In this course, 54.7% (task C1) and 62.1% (task C2) of students self-assessments were in the accurate band (between + or – 5 from marker's mark: see Table 4). This compared to 32.7% in A1, 43.8% in A2, 46.5 in B1 and 33.8% in B2.

In all the courses however, the percentage of gross over-assessors (>+10) reduced markedly from task 1 to 2 and a general movement towards more realistic assessment: in course B (a third year Accounting subject), there is a marked move towards negative self-assessment in task 2, as students received calibrating marks from task 1. In the two courses where student's self-assessment practice was best-embedded, the percentage of accurate assessors from task1 to 2 increased from 32.7% to 43.7% (Course A) and from 54.7% to 62.1% (Course C). This is neatly illustrated in Figure 2 (below). The grossest over-assessment band moderates and larger numbers of students move towards accurate and negative bands of self-assessment.

Course A exhibited the lowest self-assessment accuracy rate on the initial task (see Table 4). This could be due to the high number of recently arrived overseas students who lacked the recent calibrating experience available to the undergraduates after 2 years of study.

Figure 2: Course C (2nd year UG): Student self-assessment change - Task 1 - Task 2
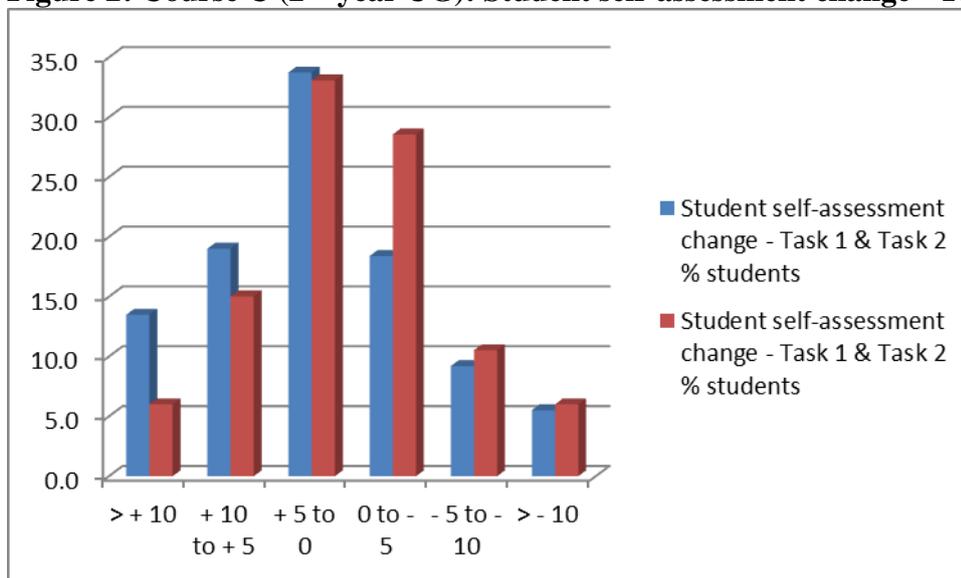
In Course C, evidence exists that the strongest under-assessors maintained a generally negative or self-critical self-assessment bias from task 1 to task 2. Students in the >-10 band generally moderated somewhat to a more realistic appraisal (see Figure 3), while still retaining an inclination to under-assess. Of the six students in the > -10 band in task 1, 5 of the 6 reduced their negative under-assessing (a common characteristic of higher-achieving students). They exhibited a marked increase in the accuracy of their self-assessment. Incidentally, all these students achieved Credit to High Distinction grades for the course.

Figure 3: Moderation of strong under-assessors task 1 – task 2
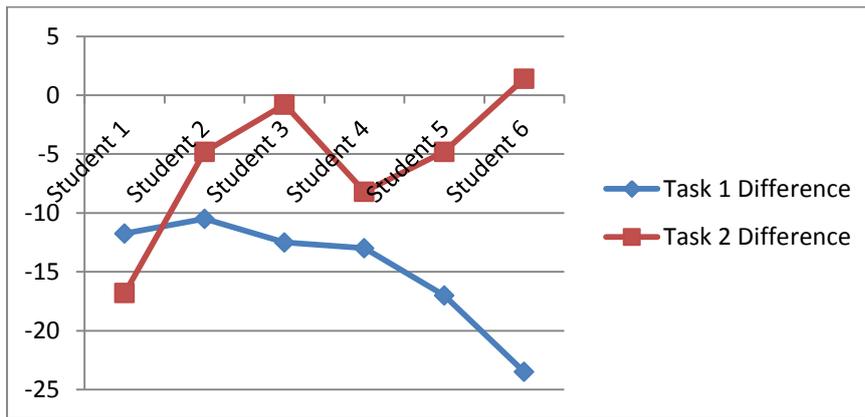
Conclusions

Opportunistic observations showed greater self-assessment accuracy in the subject courses in this study to comparable cohorts using Review for self-assessment where calibration exercises were not necessarily systematically practiced. Of the three cohorts intensively studied here, the class group incentivised with bonus marks for practicing self-assessment and accuracy in self-assessment achieved the highest participation rates and accuracy.

Overall, the increasing accuracy rates of this and the second year undergraduates indicate that, with support, practice and feedback, many students are able to quickly calibrate their overly optimistic judgments related to self-assessing against criteria. The moderation of student's most over-optimistic and most self-critical criteria assessments is affirming.

The study demonstrates both the variability of student self-assessment estimates and evidence of rapid improvement of large numbers of students towards accuracy when controlled practice and a culture of accuracy and meaningful self-assessment practice exists. The hope is that the process of frequent, thoughtful, self-assessment informs a predisposition towards seeking contextually appropriate, quality, referenced models as the basis of ongoing self-critiquing activity. The predisposition towards referencing sources of 'quality' and a reflexive habit to undertake regular evidence based (objective) self-assessment are highly desirable professional attributes for graduates.

Overall, the combined studies of attitudes towards (Carroll, 2013) and accuracy in self-assessment, highlight potential benefits for students from practicing embedded self-assessment. This leads to the conclusion that self-assessment practice should be more widely integrated into our Program (degree) designs as a powerful mechanism for developing professional predispositions and accurate judgments in our graduates.

Longitudinal research will continue into self-assessment accuracy, behavioural change and subsequent learning. Individual's continued practice of self-assessment (predisposition) and the persistence of accuracy (if any) will be a focus. As many factors impact on individuals' self-assessment accuracy, more definitive investigations into the sustained impact on learning attitudes, approaches and benefits related to self-assessment practice are planned.

Bibliography

Armor, D., Taylor, S.E., (2002). When Predictions Fail: The Dilemma of Unrealistic Optimism. In Gilovich, T., Griffin, D., & Kahneman, D., (Eds.) Heuristics and biases: The psychology of intuitive judgment. Cambridge, Cambridge University Press.

Bandura, A. (1994). Self-efficacy. Wiley Online Library.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. Annual Review of Psychology, 52(1), 1-26.

Boud, D., Lawson, R., & Thompson, D., (2013). Does student engagement in self-assessment calibrate their judgement over time? Assessment & Evaluation in Higher Education, accessed at 12/03/2014 http://www.tandfonline.com/doi/abs/10.1080/02602938.2013.769198#preview

Boud. D., (2003). Enhancing learning through self-assessment. Kogan Page, London.

Boud, D., (2000). Sustainable assessment: rethinking assessment for the learning society. Studies in continuing education 22.2: 151-167.

Carroll, D., (2013). Business student's attitudes to criteria based self-assessment and self-efficacy. Proceedings of the 30th ascilite conference, Sydney, Australia. Accessed 21/02/2014 at -http://www.ascilite.org.au/conferences/sydney13/program/proceedings.pdf

Cornelius-White, J., (2007). Learner-centered teacher-student relationships are effective: A meta-analysis. Review of Educational Research 77.1: 113-143.

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment. Psychological Science in the Public Interest, 5(3), 69-106.

Hattie, J., (2008). Visible learning: A synthesis of over 800 meta-analyses relating to achievement. Routledge.

Lew, M., Alwis, W., & Schmidt, H., (2010). Accuracy of students' self-assessment and their beliefs about its utility. Assessment & Evaluation in Higher Education, Volume 35, Issue 2.

Manitoba Education, Citizenship, and Youth. (2006). Rethinking Assessment with Purpose in Mind: assessment for learning, assessment as learning, assessment of learning. Manitoba, Canada.

Nicol, J. & Macfarlane-Dick, D., (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. Studies in Higher Education Volume 31, Issue 2.

Nicol, D., (2007). Principles of good assessment and feedback: Theory and practice. From the REAP International Online Conference on Assessment Design for Learner Responsibility, 29th-31st May, 2007. Accessed 14/02/2014 at http://ewds.strath.ac.uk/REAP07

OECD/CERI., (2008). 21st Century Learning: Research, Innovation and Policy. CERI, Centre for Educational research and Innovation. Accessed 23/02/2014 at http://www.oecd.org/site/educeri21st/40554299.pdf

Rust, C., (2003). Improving students learning by developing their understanding of criteria. Oxford Brookes University.

Ross, J. A., (2006). The reliability, validity, and utility of self-assessment. Practical Assessment, Research, and Evaluation, 11(10), 1-13.

Shepard, L., (2000). The Role of Assessment in a Learning Culture. Educational Researcher, Vol. 29, No. 7, October, pp. 4-14. American Educational Research Association. Accessed: 21/03/2014 at http://www.jstor.org/stable/1176145

Stiggins, R. J., (2002). Assessment crisis: The absence of assessment for learning. Phi Delta Kappan 83.10: 758-765.