Can students predict their grade accurately in order to self-regulate?

Regina Belski[a]; Iouri Belski[b]
La Trobe University, Melbourne, Australia[a], Royal Melbourne Institute of Technology, Melbourne Australia[b]
Corresponding Author Email: iouri.belski@rmit.edu.au

Structured abstract

BACKGROUND
Students' skills in self-regulation are critical for achieving sound learning outcomes. They are also important for effective life-long learning. It has been found that learners who actively self-regulate achieve higher grades and are more confident than their peers (Pintrich, 1995; Zimmerman & Schunk, 2001). Moreover, educational scholars are convinced that nurturing students' skills in self-regulation entails engaging them in structured, regular diagnostic assessment and self-monitoring, which leads to metacognitive reflection on their learning (Crisp, 2012; Nicol & Macfarlane-Dick, 2006).

PURPOSE
This paper investigates whether on-campus students possess adequate skills to self-regulate their learning and tries to establish the means by which we can nurture students' capacity for self-regulation.

DESIGN/METHOD
Seventy one student enrolled in a third year unit on electrical engineering were asked to predict their grades for two class tests that were conducted in weeks 6 and 9 of a 12-week semester. They evaluated the expected marks three times: (i) directly after the reading time, (ii) straight after they had completed the test and (iii) after the test solutions were presented to them. These predictions were compared with each other and with the actual test grades obtained by the students. In order to gain further insight into the results of student predictions, their Task Evaluation and Reflection Instrument for Student Self-Assessment (TERISSA) (Belski, 2007) responses were also analysed.

RESULTS
As it had been anticipated, most students were unable to make accurate predictions of their test results after reading the task. On average students over estimated their grades by nearly 26% in test 1 and by 11% in test 2. High performing students were better able to accurately predict their actual marks, then the rest of the class. Furthermore, students who actively filled in the TERISSA template for self-assessment, which engaged them in reflection performed significantly better than those who did not use TERISSA.

CONCLUSIONS
The findings of this investigation imply that students' capacity for self-regulation can be enhanced by:

- Regularly engaging them in predicting their mark.
- Helping them to regularly practice diagnostic self-assessment and reflection on their learning.
- Encouraging them to devote more time to practicing newly learnt skills and revising their theoretical knowledge.

KEYWORDS
Self-regulation, diagnostic assessment, self-assessment, deliberate practice

Introduction

The world of higher education is gradually expanding to on-line and blended learning and utilises more and more computer-based and web-based resources. As a result, direct students' engagement with teachers is shrinking and the amount of individual educational feedback from teachers is gradually replaced by automatic computer-generated feedback. These changes require educators to devote more attention to nurturing students' skills for self-regulation that are also important for the life-long learning.

Resent research has identified that many students fail to take full advantage of computer-based learning environments (CBLE) (Lee & Choi, 2011; Winters, Greene, & Costich, 2008). It has also been suggested that *"one potential mediator between the potential of CBLEs and academic performance is the quality of students' self-regulatory learning processes"* (Winters et al., 2008, p. 430).

Students' skills in self-regulation are critical for achieving sound learning outcomes. They are also important for effective life-long learning. It has been found that learners who actively self-regulate achieve higher grades and are more confident than their peers (Pintrich, 1995; Zimmerman & Schunk, 2001). Hence if we can engage student in self-regulation during on-campus, face-to-face teaching delivery, this may enable us to teach critical skills they can apply to both face to face and online learning as well as into their working life after university. Boud and Falchikov (2005) proposed a redesign of assessment in order to equip students for learning beyond university. The third proposition of the *Assessment 2020* document, developed by Boud during his ALTC Senior Fellowship, specifically conveys the need to make students more responsible learners and assessors: *"… students and teachers become responsible partners in learning and assessment*" (Boud, 2010, p. 27).

Nicol and Macfarlane-Dick (2006) proposed seven principles of good assessment practice in relation to the development of self-regulation. They focused specifically on appropriately designed formative assessment and suggested that in order *"… to develop systematically the learner's capacity for self-regulation, teachers need to create more structured opportunities for self-monitoring and the judging of progression to goals."* (p. 207). Two of their seven principles proclaim that good assessment practice *"(2) facilitates the development of self-assessment (reflection) in learning and (3) delivers high quality information to students about their learning"* (p. 205).

Crisp proposed to utilise online assessment more extensively because it can *"offer strategic and timely feedback to students, especially for large classes where individual feedback is difficult or costly in a face-to-face format*" (Crisp, 2008, p. 8). He also proposed to discriminate between four types of assessment tasks: diagnostic, formative, integrative and summative and suggested that sound diagnostic assessment can help in minimising student reliance on teachers' expectations (Crisp, 2012). Crisp recommended that *"diagnostic assessments are incorporated as an initial component in all key foundational courses and are seen as a pathway for encouraging a self-regulation paradigm in students' approaches to current and future learning"* (Crisp, 2012, p. 39).

It can be concluded that educational scholars are convinced that nurturing students' skills in self-regulation entails engaging them in structured, regular diagnostic assessment and self-monitoring, which lead to metacognitive reflection on their learning. However despite the clear need for engaging students in self-regulation, the challenges currently faced by academics are difficult to overcome without comprehending the ability of the current on-campus students to self-regulate their learning.

The ability to accurately predict their performance in various assessment activities is among the most important aspects of effective self-regulation of learning. The authors hypothesised that students would be unable to predict their marks with the accuracy that is required for effective self-regulation. In order to test the hypothesis, the authors engaged a group of engineering students enrolled in a third year unit on electrical engineering in prediction of their test performance.

Methodology

In order to judge a students' ability to predict their performance, students were asked to predict their grades for two closed-book class tests that were conducted in weeks 6 and 9 of semester 1 2013. Students made their predictions three times: i) directly after the reading time, ii) straight after they had completed the test and iii) after all the test papers were collected and the test solutions were presented to them. These student evaluations have been compared with each other and with the actual test grades.

Both tests were conducted during lecture class times and were attended by 69 and 68 students respectively (out of 71 enrolled in the unit). Each test consisted of a descriptive question that was related to new material studied in the weeks leading up to the tests. The material for the first test was covered in lectures in weeks 3 and 4 and in tutorials in weeks 4 and 5. The material for the second test was covered in lectures in weeks 5 and 6 and in tutorials in Weeks 6 to 8.

Student predictions were collected by means of the TurningPoint® clickers using the same protocol as has been previously described by the authors (Belski & Belski, 2012). Clickers were distributed to all students at the beginning of the class and were returned at the end of the class. This permitted to identify all three predictions for every individual voter and, at the same time, to ensure that individual voters could not be identified personally. Although a lecturer encouraged all students to make their predictions, voting was not compulsory and engaged 60 students during Test 1 and 59 students in Test 2.

The following are the questions that required a student response: a) at the end of the reading time: "I expect to get the following mark (out of 10) for this test"; b) after the test completion: "I think that I will get the following mark (out of 10) for this test"; c) after viewing the test solutions: "Most likely I will get the following mark (out of 10) for this test". Students had nine grade options to choose from: 10, 9, 8, 7, 6, 5, 4, 3 and '2 or less'.

The second test also incorporated the TERISSA (Task Evaluation and Reflection Instrument for Student Self-Assessment) template (Belski, 2007). TERISSA engages students in conducting two evaluations of a tasks complexity on a Likert scale from 1 (very simple) to 5 (very difficult). These evaluations are undertaken when the task is first presented and after the task has been resolved. Students then reflect on each of these evaluations and on the reasons for any discrepancy between them. Over the past nine years, TERISSA has been successfully used by many academics in tutorial classes, home and class activities, individual and group exercises, various home assignments and practical laboratory work and has resulted in notable improvements in student opinions on educational feedback (Belski, 2009, 2010).

It was expected that the analysis of the TERISSA responses would offer further insights into the results of student predictions and help to better comprehend their capability for self-regulation and the potential ways to engage students in building this capability further. The use of the TERISSA template was not compulsory. Forty four students (out of 68) filled in the TERISSA template during Test 2.

Research data

Evaluation statistics

Student predictions of their test marks for both tests as well as their actual test performance are shown in Table 1.

In the case of the first test, the students' prediction of likely marks that was made directly after reading time (Test 1, Prediction 1 in Table 1), is statistically significantly different from all other predictions they made relating to test 1 (Related Samples Wilcoxon Signed Rank Test: $p<0.002$) as well as from the actual student marks (independent samples t-test, 2-tailed).

Table 1: Predicted test performance and actual performance

	Predicted (Total)	Prediction 1		Prediction 2		Prediction 3		Actual Mark	
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Test 1	60(69)	8.52	2.12	7.30	2.77	7.12	2.72	6.86	2.97
Test 2	59(68)	7.98	2.38	7.22	2.30	6.90	2.34	7.17	2.63

No significant difference was found between student predictions after they solved the test 1 task (Prediction 2) and after the task solution was presented to them (Prediction 3). The actual marks were only statistically significantly different to the predicted marks from Prediction 1, before attempting the test ($t=3.6$, $df=127$, $p<0.001$).

Similarly, for the second test, the students' prediction of likely marks that was made directly after reading time (Test 2, Prediction 1 in Table 1), is statistically significantly different from all other evaluations (Related Samples Wilcoxon Signed Rank Test: $p<0.008$). No significant difference was found between the marks of Prediction 2 and Prediction 3. The actual marks were not statistically significantly different to any of the three Predictions (independent samples t-test, 2-tailed). It needs to be noted that if the authors hypothesised that student original prediction (Prediction 1) exceeds the actual mark, the difference between Prediction 1 and the actual test results would be statistically significant (independent samples t-test, 1-tailed: $t=1.8$, $df=124$, $p<0.05$).

As shown in Table 1, students' predictions of their own test marks in both tests went down with every successive evaluation. The actual test average mark for test 1 was lower than each of the three student evaluations. The actual test average mark for test 2 was lower than Prediction 1 and Prediction 2, but exceeded the average mark for Prediction 3.

Comparative statistics

As presented in Table 1, the average mark for Prediction 1 in test 1 (8.52 out of 10) is higher than one of test 2 (7.98) by nearly 7%. The direct comparison of these two predictions is also not appropriate because the actual marks for the tests differed by nearly 5% (6.86 versus 7.17). Moreover, the data from Table 1 show that, on average, students over evaluated (Prediction 1) their actual marks during the first test by 24% (8.52/6.86) and by only 11% (7.98/7.17) during the second test. Therefore, to make any judgements on whether students' predictions during tests 1 and 2 differed significantly, these predictions were normalised. In order to normalise the predictions, all individual prediction marks for both tests were divided by the average actual mark for the respective test and multiplied by 10 (therefore the accurate value for every prediction is 10). All normalised predictions of student performances for both tests are shown in Table 2.

Table 2: Normalised predicted test scores

	Normalised Prediction 1		Normalised Prediction 2		Normalised Prediction 3	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Test 1	12.58	2.85	10.64	4.04	10.38	3.96
Test 2	11.13	3.32	10.07	3.20	9.62	3.26

The difference between the two normalised students' predictions that occurred just after reading time (Prediction 1) is statistically significant (Independent Samples t-test, 2-tailed: $t=2.52$, $df=117$, $p<0.02$). No statistical difference was found between the normalised pairs of Predictions 2 and the pairs of normalised Predictions 3.

TERISSA statistics

Out of the 68 students, who sat the second test, 44 filled in the TERISSA template that was incorporated into the test paper. Twenty four students left the TERISSA template blank. Table 3 shows the outcomes of students' evaluation of task complexity before and after the group of 44 students solved the problem. The following tasks complexity scale was used: 1 – very simple; 2 – simple; 3 – so-so; 4 – difficult; 5 – very difficult. Table 3 also contains the actual average mark for the 44 students who filled in the TERISSA template and for the 24 students who did not use TERISSA.

Table 3: TERISSA evaluations of task complexity

	TERISSA (44)			**No TERISSA (24)**
	Evaluation		**Actual Mark**	**Actual Mark**
	B(efore)	**A**(fter)		
Mean	2.89	2.77	7.89	5.82
Std. Dev.	0.97	0.96	2.11	2.94

The difference between the original and final TERISSA evaluations presented in Table 3 was not statistically significant.

Out of 44 students who completed the TERISSA template, 13 (29.5%) evaluated the problem as more difficult before they solved it (B>A); 14 (31.8%) made the opposite judgement (B<A). The rest 17 (38.7%) did not change their opinion on task complexity (B=A). The patterns of changes of TERISSA evaluations depended on a student's actual mark. Figure 1 presents the changes in TERISSA evaluations for two groups of students: a) ones with actual marks of 8 to 10 (high performers) and b) the rest of the class, with marks of less than 8. The former group consisted of 30 students; the latter of 14. Each group is presented in Figure 1 as separate (each is out of 100%).

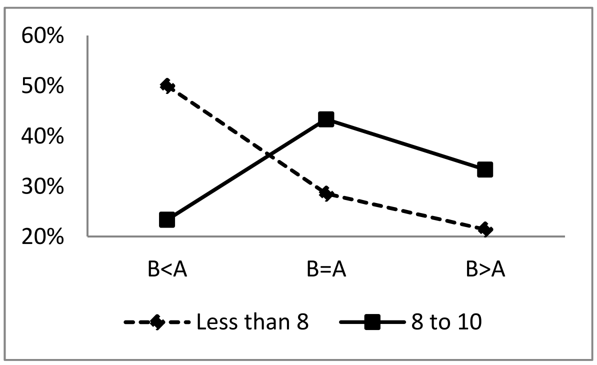

Figure 1: TERISSA evaluations: high performers and the rest of the class

There are two tendencies that are clearly present in Figure 1. Firstly, the high performers did not change their opinions on task complexity (B=A) more often than the rest of the class (44% versus 29%). Secondly, the high performers tended to over evaluate the complexity (B>A) rather than under evaluate it (33% versus 23%). The rest of the class, on the other hand, under evaluated the task complexity (B<A) more often than over evaluated it (50% versus 21%).

The actual marks of the 44 students who undertook TERISSA (TERISSA group respectively in Table 3), when compared with the marks of the rest 24 students who sat test 2, but did not fill in the TERISSA template (No TERISSA group in Table 3), were statistically significantly different between these two groups (Mann-Whitney Test, 2-tailed, Z= -3.09, p<0.002).

Discussion and conclusions

The results presented above confirm the author's hypothesis. Most students were unable to make an accurate prediction of their test results after reading the task. The other two successive predictions were more accurate. Nonetheless, the first prediction (that was made before they started solving the task) measured a student's ability to appraise their true level of knowledge and skills more accurately than the other two predictions.

It is not surprising that students were inaccurate in their predictions. It is well known that in order to undertake a proper evaluation of one's own performance a person needs to possess an appropriate level of expertise. Experts are usually capable of adequately assessing the task complexity and the time required to solve it. They also anticipate problems that could be faced while solving a task and have contingency plans in case these problems occur.

The significant difference between Prediction 1 in test 1 and the actual student performance shows that most students did not possess sufficient expertise to accurately judge their knowledge and ability and hence mark for test 1. On average students over estimated their grades by nearly 26%. It seems that the situation improved in the second test. The overall prediction for the second test was much closer to the actual performance and was only 11% above the actual. It is likely that one of the reasons of such improvement is related to the prediction activity conducted during the first test. Certainly many students noticed a significant discrepancy between their actual marks and their predictions and tried to judge their ability more accurately during test 2. This is an important step demonstrating that even a simple activity that engages students in self-assessment can lead to lasting results.

Combining the data from student prediction activities during both tests and their TERISSA reflections in test 2, it can be proposed that the group of high performers were in a much better position to accurately predict their actual marks directly after reading time. It is likely that high performers invested more time in studying the topics that were examined during the tests and, as a result, acquired more expertise than the rest of the class. Consequently, it can be inferred that more practice during learning is helpful in developing skills for self-regulation. This latter conclusion further supports a well-known fact that expertise (that improves self-regulation) can only be acquired through deliberate practice (Ericsson, 2008).

The fact that students who filled in the TERISSA template performed much better than the ones who did not can be explained in at least two ways. Firstly, it is quite possible that the students who were best prepared for the test tended to follow the lecturer's instructions more often than the ones who did not use TERISSA. Therefore, the students from this group were likely to invest time and effort in the weekly individual activities that had been suggested by the lecturer. As a result they built greater expertise and performed better. Secondly, it is possible that the students who used TERISSA during the second test had been practicing TERISSA reflection every week both during tutorials and, as it was suggested by a lecturer, individually, when solving problems outside a class. These activities may have improved the students' ability to assess their own learning difficulties and may have resulted in deliberate practice (Belski, 2010).

The drop in the expected mark from Prediction 1 to Prediction 2 can also be explained by lack of expertise, and further supports the abovementioned conclusion. While solving the task, students were likely to experience unexpected problems that they did not anticipate while predicting their mark after reading time. In line with the reduction of expected mark from Prediction 1 to Prediction 2, the last prediction that happened after the solution was revealed (Prediction 3) reduced students' predictions even further. Most likely some students discovered that they made some silly mistakes (e.g. mathematical) or even realised that their solutions were entirely wrong (for test 1 seven students dropped their predicted marks from Prediction 1 to Prediction 3 by 5 marks or more; five students recorded the same drop in their predictions in test 2).

The results of this investigation demonstrate that even after two years of study students were:

a) unable to accurately predict their test performance after reading time;
b) students' prediction improved in the second test, after they practiced predicting their mark in test 1;
c) students who performed better were more accurate in predicting their performance compared with students who performed poorly;
d) students who followed the TERISSA procedure outperformed students who did not use TERISSA.

The above-mentioned findings imply that students' capacity for self-regulation can be enhanced by:

- Regularly engaging them in predicting their mark. This can be achieved by simply embedding a prediction template into a test/assignment paper.
- Helping them to regularly practice diagnostic self-assessment and reflection on their learning. This can be achieved in numerous ways (see www.assessmentfutures.com for the possible techniques and approaches).
- Encouraging them to devote more time to practicing newly learnt skills and revising their theoretical knowledge.

Acknowledgements

The authors wish to thank the reviewers for their valuable comments and helpful suggestions.

References

Belski, I. (2007). *Using task evaluation and reflection instrument for student self-assessment (terissa) to improve educational assessment and feedback.* Paper presented at the Proceedings of the 18th Conference of the Australasian Association of Engineering Education (AaeE), Melbourne, Australia.

Belski, I. (2009). *Improving student satisfaction with feedback by engaging them in self-assessment and reflection*. Paper presented at the ATN Assessment Conference, Melbourne, Australia.

Belski, I. (2010). *The impact of self-assessment and reflection on student learning outcomes.* Paper presented at the 21st Annual Conference for the Australasian Association for Engineering Education, Sydney, Australia.

Belski, I., & Belski, R. (2012). *Clickers: Enabling valuable diagnostic feedback to both students and instructors in real time.* Paper presented at the IASTED International Conference: Computers and Advanced Technology in Education (CATE 2012), Napoli, Italy.

Boud, D. (2010). Student assessment for learning in and after courses: Final report for senior fellowship (pp. 32). Sydney: UTS.

Boud, D., & Falchikov, N. (2005). Redesigning assessment for learning beyond higher education. In A. a. A. Brew, C. (Ed.), *Research and development in higher education* (pp. 34-41). Sydney: HERDSA.

Crisp, G. (2008). Raising the profile of diagnostic, formative and summative e-assessments. Providing e-assessment design principles and disciplinary examples for higher education academic staff. *Australian Learning and Teaching Council Associate Fellowship Report. Retrieved January 2012 (http://www.olt.gov.au/resources?text=crisp)*, 48p.

Crisp, G. (2012). Integrative assessment: Reframing assessment practice for current and future learning. *Assessment & Evaluation in Higher Education, 37*(1), 33-43. doi: 10.1080/02602938.2010.494234

Ericsson, K.A. (2008). Deliberate practice and acquisition of expert performance: A general overview. *Academic Emergency Medicine, 15*(11), 988-994.

Lee, Y., & Choi, J. (2011). A review of online course dropout research: Implications for practice and future research. *Educational Technology Research and Development, 59*(5), 593-618.

Nicol, D.J., & Macfarlane-Dick, D. (2006). Formative assessment and selfregulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199–218. doi: 10.1080/03075070600572090

Pintrich, P.R. (1995). *Understanding self-regulated learning*. San Francisco, CA: Jossey-Bass.

Winters, F.I., Greene, J.A., & Costich, C.M. (2008). Self-regulation of learning within computer-based learning environments: A critical analysis. *Educational Psychology Review, 20*(4), 429-444.

Zimmerman, B.J., & Schunk, D.H. (2001). *Self-regulated learning and academic achievement: Theoretical perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates.

Copyright statement