Prediction and Post-diction of Performance in Class: A Self-assessment Study

Kawtar Tani
School of Business and Computing, UCOL, Private Bag 11022
Palmerston North 4442, New Zealand

[Abstract] This study investigated the accuracy of the self-assessment ability of students enrolled in tertiary education, pre-examination and post-examination. The participants included 174 undergraduate students enrolled in different programs in a tertiary education institution. Students' self-assessment abilities were measured by asking students to assess their performance prior to and after written examinations. The relationship between students' self-assessment and performance were analyzed to determine the accuracy of the self-assessment. There was strong statistical evidence that predictions and post dictions of exam scores were more optimistic than scores achieved in exams. Students with above- average academic abilities were much better at predicting their performance on exams, while students with lower academic abilities were more likely to overrate their abilities. The latter were also significantly optimistic; there was no evidence that students who scored above the median in the final exam exhibited optimism in their predictions made either before or after the exams.

[Keywords] assessment, learning strategies, test performance, prediction, study skills

Introduction

In education, institutions often predict exam scores of new entrants using statistical methods or based on teachers' predictions. However, The Cambridge Assessment study in the United Kingdom reported that based on an analysis of almost 190,000 papers by its exam board, just 48.3 percent of teachers' predictions made in 2012 were proved correct when final results were published. The accuracy level had dropped from 54.65 percent when the process was last carried out in 2009. Because of such erroneous predictions, thousands of students were at risk of missing out on university places (*The Telegraph*, 2013). Other studies reported that elementary exam results can be predicted by measuring the length of fingers (Brosnan, 2008).

Most predictions are subject to two kinds of error (Wilkins, 1980): (1) the individual who is predicted to fail may succeed, or (2) the individual who is predicted to succeed may fail. The first kind of error is usually termed "over-prediction," and the second error is termed "under-prediction." These errors apply to all decisions or estimates irrespective of the means by which such decisions or estimates are derived.

The accuracy of test prediction, prior to an examination, relies on the assessor's knowledge of the topic to be assessed. It can be said that accurate self-assessment pre-examination can help students be better prepared for their examinations by reducing the occurrences of false positives and false negatives.

One previous study (Hacker, Bol, Horgan, & Rakow, 2000) looked at test prediction and performance in a classroom context. The authors examined the relationship between the accuracy of test predictions and performances and reported that higher-performing students made accurate predictions than low-performing students. The authors, however, did not examine whether or not and how predictions can be used to self-regulate student learning.

The purpose of this study was to evaluate the accuracy of the self-assessment ability of students enrolled in tertiary education, pre- and post-examination, and determine how the

information from summative self-assessments can be used formatively by students or faculty to guide their efforts and activities in subsequent courses. Students' self-assessment ability was measured by 1) asking students to assess their performance prior to written examinations, and 2) asking them to estimate their performance following the written examination. The relationship between students' self-assessment and performance was analyzed to determine the accuracy of the self-assessment.

Method

Participants

There were 174 students (81 females and 92 males) enrolled in a range of courses at a tertiary education institution in New Zealand during the second semester of 2015 and the first semester of 2016. Of these students, 64 were on year one of their study, 65 on year 2, and 45 on year 3. There were 147 domestic students and 27 international students. Participants defined their ethnicity as New Zealand European (120), Indian (27), Maori (15), Pacifika (6), or Other (6). The mean age of participants was 26.5yr (SD = 7.21, range 19 to 49).

Materials

Prospective participants were given a detailed account of the research in class and were asked to participate in the study. Participation was voluntary, and no rewards were given for participating. Research approval was obtained from the institution in which the research was conducted.

Procedure

Pre-examination grades were collected from participants in class during exam week reviews, normally one week prior to written examination. Post-examination data were collected at the end of the written examination via a form attached to the exam scripts requesting those who wished to participate to grade their performance on the written text they had just completed. Students were also asked to provide their student ID numbers on the attached form. The student ID numbers were used to derive other data, including age, level of study (year 1, 2 or 3), and type of enrolment (domestic vs. international). Data were analyzed using SPSS version 22. All tests were conducted as two-tailed, with alpha set *a priori* at .05.

Results

There was a small amount of missing data. However, as it appeared randomly distributed across participants and variables, no participants were excluded from the analyses that follow; where responses are missing, it is reflected by either *df* or *n*. There were 174 participants of whom 92 were male and 81 were female (one participant did not provide information about gender), a difference which was not statistically significant, $X^2(1) = .70$, $p = .40$.

To explore the accuracy of students' predictions of their exam scores, their estimated scores predicted *before* or *after* the exam were subtracted from actual scores. Positive scores would therefore suggest pessimistic scores, whereas negative scores would suggest optimistic scores. Overall, the difference between actual and scores predicted directly before the exam was -5.48 (*SD* = 18.97, range -72.22 to 32.86); the difference between actual and scores predicted directly after the exam was -3.46 (*SD* = 13.06, range = -55.56 to 32.86). One-sample *t*-test (test value = 0) indicated that both before and after predictions were statistically significantly optimistic: $t(173) = 4.73$, $p < .001$, $d = .29$ (before) and $t(170) = 4.51$, $p < .001$, $d = .26$ (after). However, predictions

made before the exam were significantly more optimistic than those made after the exam, $t(172)$, = 3.06, $p = .003$, $d = .182$. There was no evidence that predictions varied according to participants' gender, *ns*.

Correlations (Pearson's *r*) were statistically significant between all three pairs: before vs. after prediction $r = .589$, $n = 172$, $p < .001$; before vs. actual prediction, $r = .278$, $n = 174$, $p < .001$; and after vs. actual predication, $r = .673$, $n = 172$, $p < .001$. The before vs. after ($Z = 3.60$, $p < .001$) and after vs. actual ($Z = 4.89$, $p < .001$) correlations were both significantly stronger than the before vs. actual correlation; however, the before vs. after and the after vs. actual correlations did not differ significantly in size, ($Z = 1.29$, $p = .09$).

There was no evidence of differences by participants' year of study (1, 2, 3) on the difference between actual exam scores and predictions made before the exam, $F(2, 173) = 2.46$, $p = .09$, $\eta^2 = .028$; however, there was evidence of a significant difference by participants' year of study (1, 2, 3) on the difference between actual exam scores and predictions made after the test, $F(2, 171) = 3.28$, $p = .04$, $\eta^2 = .037$ (year 1 M = -4.11, SD = 12.42, year 2 M = -6.88, SD = 10.07, year 3 M = -.68, SD = 15.07). LSD test indicated a significant difference of 6.21, $p < .05$ between groups 2 and 3, but no other pairwise differences.

To further investigate the accuracy of participants' predictions by actual performance in the exam, a median split on final exam scores was conducted. As the median exam score was 60%, this resulted in $n = 84$ participants scoring below the median and $n = 90$ scoring 60% or above. There was statistically significant evidence of a difference in the accuracy of predictions made before the exam, $t(172) = 7.03$, $p < .001$, $d = 1.07$ (mean difference score for those above the median = 1.84, $SD = 14.21$, mean difference score for those below the median = -16 .09, $SD = 19.21$) and, also, there was significant evidence of a difference in the accuracy of predictions made after the exam, Welch's $t(146.52) = 7.23$, $p < .001$, $d = 1.13$ (mean difference score for those above the median = 1.62, $SD = 8.97$, mean difference score for those below the median = -10.6, $SD = 12.69$). (As Levene's test for equality of variances was significant, we reported Welch's *t*-test with adjusted *df*.) That is, those scoring below the median in the actual exam were significantly more optimistic than those scoring above the median.

One sample *t*-test, using a test score of zero, and difference scores of before and after predictions as the dependent variable suggested that those who scored above the median (≥60%) on the actual exam were neither optimistic nor pessimistic in their predictions made directly before the exam, $t(83) = 1.28$, $p = .223$, $d = 0.13$, or directly after the exam, $t(82) = 1.70$, $p < .093$, $d = 0.18$; however, participants who scored below the median were significantly optimistic in both their predictions made both before, $t(89) = 7.67$, $p < .001$, $d = 0.84$ and after the exam, $t(88) = 7.64$, $p < .001$, $d = 0.84$.

Discussion

There was strong statistical evidence that predictions about exam scores made directly before the exam and directly after the exam (post-dictions) were more optimistic than scores actually achieved in the exam. This result is consistent with findings reported earlier by Mueller (1970) and by Balch (1992). However, the omnibus finding reported above conceals a much more interesting finding: only those students scoring below the median in the actual exam were significantly optimistic; there was no evidence that students who scored above the median in the final exam exhibited optimism in their predictions made either before or after the exam. Thus, students with above average academic abilities, as measured by their performance in a single exam, were also much better at predicting their performance in exams, while students with lower

academic abilities were more likely to overrate their ability. The predictive accuracy, by actual achievement, is consistent with much earlier evidence (e.g. Shaughnessy, 1979), where it was reported that good students are more accurate in predicting the correctness of their answers to individual test items compared with poorer students.

Interestingly, the significant optimism that poorer students displayed before the exam was present despite the use of measures intended to enhance students' accuracy of self-assessment. For example, students' pre-exam predictions were recorded towards the end of exam review sessions conducted the week before the exams, in which lecturers discussed with students their preparedness for the upcoming exams.

There is potentially a serious problem with overly optimistic predictions of subsequent exam performance; that is, immediate feedback may affect the learning and motivation of students (Skinner, 1958). Thus, students who performed below average concluded that they were well prepared when perhaps they were not and consequently may have relaxed their preparation for exams. Contrarily, students who performed above the median in the actual exam evaluated themselves less confidently when compared to those scoring below the median and may then have felt that further exam preparation was required.

An interesting result was the optimistic postdiction of performance made by students at year 2; these students appeared to be more optimistic in predicting their post-exam scores than students in year 3. One reason could be by year 3, students are expected to be more realistic about their own skills and abilities and to be better at self-assessments than in years 1 and 2. Year 1, being an introductory level to tertiary education , students are not required to meet high academic expectations; as a result, students may carry over these low expectations to year 2 and over-estimate their academic performances, which is consistent with the results of this study.

There are a number of limitations in the current study. First, this was a correlational rather than an experimental study. As such, it is not possible to conclude whether students who performed above average in the exam are better at predicting exam performance, or whether predictions made before an exam encourage preparation for the exam. For students who performed below the median in the exam, the opposite also cannot be known. Second, the exams used in this study were essay-type studies. Not using multiple choice tests can be subject to subjective scoring by the instructor.

Future research might seek to explore the extent to which predicted performance affects revision strategies of students preparing for exams, although, for ethical and pragmatic reasons, this would be unlikely to be an experimental design.

Conclusion

Despite the advancements in teaching practices and strategies, tools, and techniques, the findings of this study were strikingly consistent with a study conducted by Balch (1992). Balch reported an overestimation of below-average students' preparation for and performance on tests and an accuracy of predictions of students above average. This trend in unrealistic beliefs of below-average students suggests that poorer students are unlikely to improve their study habits, which is costly for all concerned (Balch, 1992).

This study showed how self-assessment can be used as a tool to enhance students' learning skills and motivate them towards better study habits. High-performing students are already realistic enough to recognize their academic capabilities and judge their academic performances. Thus, instructors can focus on helping poorer students improve their self-assessment skills.

One explanation for poorer students' optimism is that these students lack understating of the material and, instead, rely on memorizing material for the exam. The lack of comprehension of

material may contribute to overestimation of exam readiness and exam performance, as shown in this study. If this explanation is valid, instructors might best focus on helping poorer students with comprehension of material before attempting to teach them self-assessment skills.

References

Balch. R. W. (1992). Effect of Class Standing on Students' Predictions of Their Final Exam Scores. *Teaching of Psychology, 19*(3), October 1992.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 1*(1), 671-684.

Hacker, D., Bol, L., Horgan, D., & Rakow, E. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology, 92*(1), 160-170.

Hyde, T. S., & Jenkins, J. J. (1969). Differential effects of incidental tasks on the organization of recall of a list of highly associated words. *Journal of Experimental Psychology, 82*, 472-481.

Hyde, T. S., & Jenkins,J. J. (1973). Recall for words as a function of semantic, graphic, and syntactic orienting tasks. Journal of Verbal Learning and Verbal Behavior, 12, 471-480.

Mueller, R. H. (1970). Is self-grading the answer? *Journal of Higher Education, 41*, 221-224.

Shaughnessy, J. J. (1979). Confidence-judgment accuracy as a predictor of test performance. *Journal of Research in Personality, 13*, 505-514.

The Telegraph (2013). Exam chiefs: most predicted A-level grades are wrong. Retrieved from: http://www.telegraph.co.uk/education/educationnews/10397499/Exam-chiefs-most-predicted-A-level-grades-are-wrong.html

Tousignant, M., & DesMarchais, J. E. (2002). Accuracy of Student Self-Assessment Ability Compared to Their Own Performance in a Problem-Based Learning Medical Program: A Correlation Study. *Advances in Health Sciences Education, 7*(1), 19-27.

Brosnan. M. J. (2008). Digit ratio as an indicator of numeracy relative to literacy in 7-year-old British schoolchildren. *The British Journal of Psychology, 99*(1), 75-85.

Wilkins, L. (1980). Problems with existing prediction studies and future research needs. *The Journal of Criminal Law & Criminology, 71*(2).