

A Preliminary Study of Grade Forecasting by Students

Michael J Armstrong, Associate Professor

Finance, Operations, & Information Systems, Faculty of Business, Brock University

St Catharines, ON, L2S 3A1, Canada, michael.armstrong@brocku.ca

ABSTRACT

This experiment enabled undergraduate business students to better assess their progress in a course by quantitatively forecasting their own end-of-course grades. This innovation provided them with predictive feedback in addition to the outcome feedback they were already receiving. A total of 144 students forecast their grades using an instructor-prepared spreadsheet, and then responded to a brief survey. Of these participants, 29% said the forecast grades were lower than expected, while 6% said they were higher. Subsequent to the forecast, 47% of the respondents said they were studying more than planned, while 3% said they were studying less. The relative difference between the students' forecast grades and their prior expectations showed no direct influence on subsequent motivation or studying effort. Instead, increased studying was reported by students who had experienced increased anxiety, increased motivation, or positive impressions subsequent to the forecasting experience, as well as by students who had received low absolute grade forecasts.

Subject Areas: feedback, forecasting, self-regulated learning, business education.

This is the peer reviewed version of the following article: [Armstrong MJ, 2013. "A preliminary study of grade forecasting for students", Decision Sciences Journal of Innovative Education 11 #2, 193-210.

DOI: 10.1111/dsji.12003], which has been published in final form at

<http://onlinelibrary.wiley.com/doi/10.1111/dsji.12003/abstract>. This article may be used for non-commercial purposes in accordance with [Wiley Terms and Conditions for Self-Archiving](#).

INTRODUCTION

Feedback and self-assessment are important parts of the learning process. For example, the *self-regulated learning* (SRL) model theorizes that students establish and pursue their own learning goals (e.g., getting a “B” in the course). As their studies proceed, they compare their actual progress to their goals, and take action (e.g., studying) to ensure they achieve those goals (Butler & Winne, 1995; Nicol & Macfarlane-Dick, 2006)). In this dynamic process, students periodically assess their own learning, in part by interpreting *outcome feedback* (e.g. test marks) they receive from their instructors. This kind of purposeful feedback loop in learning theory is similar to ones seen in other contexts, such as statistical process control in manufacturing.

Unfortunately, even if students follow this process, their self-assessments may only approximate their actual learning (Nowell & Alston, 2007; Sitzmann, Ely, Brown, & Bauer, 2010). Students tend to over-estimate their learning progress, with the weakest ones being the most optimistic (Kruger & Dunning, 1999). This can hinder their academic success, as they have no reason to change their behavior if they believe they are already doing well.

Instructors can reduce this problem by providing additional forms of feedback. Strasser (2003), for example, supplied students with practice exams to help them assess their learning. Students compared their answers to the solution set, and also asked the professor to explain the reasoning behind the solutions. Likewise, Pavlovich, Collins and Jones (2009) had students keep written journals in which they reflected on their learning as the semester progressed.

This paper describes an initial empirical study of another form of supplementary feedback. Students in an undergraduate business course were given the opportunity to use their term marks to quantitatively forecast their own final mark while the course was still underway. This innovation provided participants with forward-looking *prognostic* feedback (e.g., predicted

future course grades) in addition to more conventional backward-looking *diagnostic* feedback (e.g., actual past quiz marks), so as to help them better assess their own performance.

This seems to be the first study of quantitative grade forecasting by students for their own use. Some existing research has students “predicting” their own course grades, in the sense of simply asking them what grade they expect to receive, without providing any quantitative tools (e.g., Nowell & Alston, 2007). Other studies involve quantitative forecasts of student success, but for use by educators. For example, Beck and Davidson (2001) used academic orientation surveys and high school marks to predict first-year university grades, so that counselors could identify the students at greatest risk of failing. But there seems to be little research where the students themselves receive the forecasts. Beyer (1971) performed an experiment in which students entering university were given forecasts of their first-semester grade point averages. Half of these forecasts were biased upward, to investigate whether that manipulation would affect student performance. No significant impact on end-of-semester grades was found.

The research herein has three main goals. The first is to conduct an exploratory study on grade forecasting as a form of supplementary feedback. Is grade forecasting of interest to the students? If they try it, will they respond to it in any meaningful way, or simply ignore it?

The secondary goal of this research is to test some competing propositions from learning theory in a novel context. SRL suggests that students should respond to a grade forecast in terms of the *relative* differences between the forecast and their prior expectations (as will be described in Hypotheses 3 and 4 in the next section). By contrast, other arguments suggest that students should instead respond to the *absolute* forecasts (Hypotheses 5 and 6).

The tertiary goal of this study is to verify whether previous results about student learning still hold in this new setting. This includes the tendency for students to be overly optimistic (Hypothesis 2) and for stronger students to be more active in course activities (Hypothesis 1).

The research herein was inspired by a parallel problem in business, where managers often have trouble making good sales forecasts using only their own judgment. Their ability improves, however, if they use statistical forecasting techniques (Leitner & Leopold-Wildburger, 2011) and receive feedback on their accuracy (Goodwin, Önkal, & Thomson, 2010).

Comparable benefits from forecasting could apply in the university learning context. *Cognitive feedback* (e.g., recommended study practices) can help students become more self-aware (Feys, Anseel, & Wille, 2011). *Functional validity feedback* that compares the learner's estimates of achievement to their actual performance is especially relevant. It enables students to better calibrate their self-assessments (Butler & Winne, 1995) and strengthen the correspondence between self-assessed and actual learning (Sitzmann, Ely, Brown, & Bauer, 2010).

HYPOTHESES

Student Participation

Student performance is known to be associated with many kinds of student engagement, such as the frequencies with which they attend classes (Westerman, Perez-Batres, Coffey, & Pouder, 2011), or view homework solutions online (Klassen & Biktimirov, 2007). There are several explanations for this link. Learners with stronger self-regulation tend to monitor their progress more carefully and seek out more feedback (Butler & Winne 1995). Weaker students might avoid activities where that weakness could be revealed. And students who are involved in jobs or other extra-curricular activities would simply have less time left for academic activities. Thus in this study, higher-performing students may be more likely to voluntarily try grade forecasting.

Hypothesis 1: Higher forecast marks will be associated with higher rates of trial of the forecasting tool.

Excessive Optimism

As already noted, learner self-assessments tend to be overly optimistic (Kruger & Dunning, 1999; Nowell & Alston, 2007). Students with the least progress seem to overestimate the most, while stronger students may underestimate their learning (Kruger & Dunning, 1999). These problems could arise if, for example, learners are not adequately *monitoring* the information (e.g., test marks) they receive; they may observe the relevant cues, but misperceive what those cues predict (Butler & Winne, 1995). If students have unreasonable grade expectations, they should find those expectations differing significantly from the quantitative forecasts.

Hypothesis 2a: Forecast grades will tend to be lower than students expect (a negative grade difference), rather than higher (a positive grade difference).

Hypothesis 2b: Higher forecast grades will be associated with positive grade differences, while lower forecast grades will be associated with negative grade differences.

Responding to Relative Differences

Inherent to SRL is the theory that students should respond to differences between their learning goals and their actual progress, and act to reduce those differences (Carver & Scheier, 1990; Nicol & Macfarlane-Dick, 2006; Sitzmann, Ely, Brown, & Bauer, 2010). If students see that their forecast grades are different than expected, they presumably will compensate by adjusting their studying efforts. Of course, this response to feedback is a theoretical proposition; whether any given student actually behaves that way in practice could depend on factors such as their motivation, learning skills, and past experience.

Hypothesis 3: Positive grade differences will be associated with decreased studying, while negative grade differences will be associated with increased studying.

In addition to this behavioral response, students may also experience cognitive and affective (emotional) responses to any difference between expected and actual performance. For example, students who discover their progress is less than expected may generate negative affect and lose motivation. Those with the most unrealistic beliefs may become the most discouraged (Carver & Scheier, 1990). As well, student self-assessments can be closely related to affective outcomes of learning (Sitzmann, Ely, Brown, & Bauer, 2010). So if grade forecasts reveal that those self-assessments need changing, the associated affective outcomes may also change.

Hypothesis 4a: Positive grade differences will be associated with positive student feelings about the forecasting exercise, while negative grade differences will be associated with negative feelings.

Hypothesis 4b: Positive grade differences will be associated with positive student recommendations about the exercise; while negative grade differences will be associated with negative recommendations.

Hypothesis 4c: Positive grade differences will be associated with decreased student motivation toward the course, while negative grade differences will be associated with increased motivation.

Responding to Absolute Forecasts

Some students instead might respond to the value of the forecast grade itself, regardless of whether it differs from their self-assessments.

For example, expectancy theory suggests that students' motivation depends on their beliefs about the links between their efforts and their course outcomes. If students receive

information that makes them believe these links are strong, then that may result in increased motivation and effort. Feys, Anseel, and Wille (2011) found that high feedback scores (e.g., “excellent”) led to more favorable emotional responses (e.g., happy, motivated) among trainees, while low scores (e.g., “extremely weak”) led to more unfavorable ones. They also reported that high scores led to more participation in later developmental activities. Even simple words of encouragement might sometimes be enough to increase student performance (Van Oudenhoven, Siero, Veen, & Siero, 1982). This suggests that students will respond positively to high forecast grades, and negatively to low forecast grades.

Alternatively, this correlation might instead be negative. For some students, their grade goal might not be a numerical target to hit, but rather a hurdle to clear: they simply want to pass the course. In those cases, it would be low forecasts that lead to increased studying, to ensure clearing the hurdle; high forecasts would lead to reduced studying. The following hypotheses suppose that the relationship between forecasts and responses is positive.

Hypothesis 5: High forecasts will be associated with increased studying, while low forecasts will be associated with decreased studying.

Hypothesis 6a: High forecasts will be associated with positive feelings, while low forecasts will be associated with negative feelings.

Hypothesis 6b: High forecasts will be associated with positive recommendations, while low forecasts will be associated with negative recommendations.

Hypothesis 6c: High forecasts will be associated with increased motivation, while low forecasts will be associated with decreased motivation.

Interrelated Outcomes

Learners interpret feedback according to their beliefs about subjects, processes, and outcomes. (Butler & Winne, 1995; Nicol & Macfarlane-Dick, 2006). Their behavioral response may not be just a direct outcome of the feedback itself, but also a second-order outcome triggered by their affective and cognitive responses. Thus in the present experiment, changes in student behavior may be driven in part by their feelings and thoughts about the forecasting exercise itself.

Hypothesis 7a: Positive feelings will be associated with increased studying, while negative feelings will be associated with decreased studying.

Hypothesis 7b: Positive recommendations will be associated with increased studying, while negative recommendations will be associated with decreased studying.

Hypothesis 7c: Increased motivation will be associated with increased studying, while decreased motivation will be associated with decreased studying.

It is also reasonable to think that there will be correlations among these responses. Learner reactions to training can influence their motivation and self-efficacy, as well as their knowledge (Sitzmann, Brown, Casper, Ely, and Zimmerman, 2008). In the present study, students who respond along one dimension are likely to respond similarly along others.

Hypothesis 8a: Positive feelings will be associated with positive recommendations, while negative feelings will be associated with negative recommendations.

Hypothesis 8b: Increased motivation will be associated with positive recommendations, while decreased motivation will be associated with negative recommendations.

Hypothesis 8c: Positive feelings will be associated with increased motivation, while negative feelings will be associated with decreased motivation.

Overall Path Model

Combining Hypotheses 2 to 8 leads to a path model with 15 links (Figure 1). The logical flow is from left to right: the forecast grades influence the grade differences; the forecasts and differences influence student feelings, recommendations and motivation; and all of these influence their studying efforts.

Note that this network need not turn out to be all true or all false. In particular, Hypotheses 3 and 4 are competing with 5 and 6. The former two assert that the *relative* forecast grade (i.e., the difference between expectation and forecast) will be the primary driver of student responses, whereas the latter two predict that the *absolute* forecast grade will play this role.

METHOD

Student Population

This study took place in the introductory operations management course of a mid-sized Canadian university accredited by the Association to Advance Collegiate Schools of Business. This course is required for all second-year undergraduate business students, of which there were 616 in 2011-2012. Their preceding studies include one course in probability and statistics, and one in data analysis; the latter includes an introduction to business forecasting. The students are about 2/3 male and 1/3 female, about 95% domestic and 5% foreign, and have a median age of 19 years.

Data Collection Procedure

At the beginning of each semester, the grades from the previous offering of the course were analyzed using linear regression. This produced formulas for point and interval estimates of the end-of-course grades, based on the marks from tests and assignments.

For example, a regression was run using the marks from the September-December 2010 offering of the course. Those marks consisted of 5 components: term test 1 was marked out of a possible 30 points, and counted for 15% of the overall course grade; test 2 was marked out of 28

and counted for 15% of the course grade; assignment 1 was marked out of 20 and counted for 10% of the course grade; assignment 2 was marked out of 20 and counted for 10% of the course grade; and the final exam was marked out of 100 and counted for 50% of the course grade. The analysis showed that in this case only the test marks were statistically significant predictors of the exam mark. The best-fitting regression equation for the exam mark was $Exam = 18.04 + 1.135 Test1 + 1.237 Test2$. This equation was then used for forecasting the exam marks for the January-April 2011 offering of the course. The predicted exam marks were combined with the actual test and assignment marks to forecast the overall course grade as $Grade^* = 0.5 Exam^* + 0.5 Test1 + 0.5357 Test2 + 0.5 Assign1 + 0.5 Assign2$, as per the course syllabus. These equations were embedded within a spreadsheet along with instructions regarding their use, as shown in Appendix 1. (The Excel spreadsheet file is available from the author upon request.)

The forecasting spreadsheet was posted on the course website and demonstrated in class at the beginning of week 10 of the 12-week course; this was about 3.5 weeks before the exam. Each student was free to try the spreadsheet on their own; its use was optional and anonymous. In week 12 a brief survey was distributed in class regarding the forecasting experiment; see Appendix 2. The survey also was optional and anonymous.

During 3 semesters from January 2011 to April 2012, this procedure was implemented in 10 sections of the course: 5 taught by the author, and 5 taught by 3 other professors. These sections had a combined enrollment of 465 students from across all business specialties (accounting, finance, marketing, etc.). This represented about three-quarters of the 616 second-year business-student population at this university. The 4 professors all used the same textbook, but developed their own lectures, assignments, and tests; consequently, slightly different forecasting formulas were used in each section of the course. In each case the final exam was

cumulative (i.e., it tested material from the entire course, including content previously covered by the term tests) and counted for approximately half of the overall course grade. This diversification across time periods and instructors increased the sample size and improved the generalizability of the results.

Instead of posting the spreadsheets online for student use, the instructors could have forecast the grades themselves. However, involving the students in the forecasting procedure made it more transparent to them, and allowed them to exercise their self-assessment skills (Nicol & Macfarlane-Dick, 2006). As well, students can be more accepting of feedback that is computer-moderated, rather than instructor-provided (Kluger & Adler, 1993).

Sample Size

Out of 465 students enrolled in the course, 282 (60.6%) completed the survey. Of these, 138 (49%) indicated they had not tried the forecasting tool. When asked why not, 51 (37%) said they already could estimate their own grade, 23 (17%) said they were not interested, 11 (8%) said they preferred not to know their mark, and 53 (38%) indicated it was for some other reason.

The other 144 students (51%) said they had tried the forecasting tool, and so answered the remaining questions. These are shown in Appendix 2, where the figures along the left side indicate the number and percentage of students choosing each answer. These data were analyzed with Minitab 14 statistics software and Amos 20 structural equation modeling (SEM) software.

RESULTS

Accuracy of the Forecasts

For the 465 students who completed the course during this study, the average actual grade was 67.0 out of 100. By comparison, the spreadsheet formulas forecast an average grade of 67.8, giving a mean forecast error of +0.8 marks. The standard deviation of these forecast errors is

7.35. The mean absolute deviation is 5.84; that is, the actual grades for individual students were higher or lower than their forecast grades by an average of 5.84 marks.

Student Participation

The survey responses indicate that the proportion of respondents who tried the forecasting spreadsheet varied with student performance. Table 1 displays the number and percentage of survey respondents who tried the spreadsheet, broken down by grade category; it also gives the comparable breakdown for all students enrolled in the course. For example, 30 students with forecast grades of 80-100 tried the spreadsheet; this was 21% of the 144 total. By comparison, there were 57 students in the course with forecast grades in that range, which was 12% of the 465 enrolled. The numbers indicate that the highest performing students were 7.6 times more likely than the lowest-performing students to report forecasting their grade. A chi-squared test comparing the sample distribution against the population gives $p < .001$, confirming that the difference in reported trials is statistically significant.

Responses to Individual Survey Questions

About 29% of the students reported that their forecast grade was noticeably lower than expected, whereas only 6% reported it being higher. A test using the binomial distribution confirmed that this imbalance is statistically significant ($p < .001$). Regarding their behavior subsequent to the forecast, 47% of the respondents said they were studying more than planned, while 3% said they were studying less ($p < .001$).

Most students experienced notable emotional reactions to their forecast, with 31% reporting positive feelings and 35% reporting negative ones. These proportions are not significantly different from each other ($p = .538$). Their reported changes in motivation towards the course tended to be positive, with 56% saying that their motivation increased while only 7%

said it decreased ($p < .001$). With respect to future use of forecasting, 74% of respondents recommended in favor of it, while 6% recommended against ($p < .001$).

Relationships: Chi-squared Tests of Association

After examining the responses to individual survey items, the analysis next looked for potential relationships between the items. The first method used for this purpose was to organize the responses for each of the 15 hypothesized pairings into a 2-way contingency table. A chi-squared test was then applied to each table to confirm whether there was any significant association between the two survey items.

For example, consider the potential association between students' feelings and their studying. Feelings had 3 possible responses (negative, neutral, or positive), as did studying (less, no change, or more), so there were 9 possible combinations spread across a 3 x 3 table. Since there were less than 5 responses in each of the "studying less" cells, those counts were combined with their counterparts from the "no change" response to end up with a 2 x 3 table. The calculated value for Pearson's chi-squared statistic was statistically significant ($\chi^2 = 8.381$, $df = 2$, $p = .015$), indicating an association between student feelings and studying effort.

Table 2 displays the calculated p -values for the chi-squared tests of all 15 possible associations. Of these, 10 show evidence of significant associations ($p \leq .05$). The forecast grade is associated with the grade difference, and with student feelings and studying. The grade difference is associated with student feelings and recommendations. Student feelings and motivations are both associated with their recommendations and studying. Finally, their recommendation is associated with their studying as well.

Relationships: Structural Equation Modeling

The second method used to identify relationships among the survey items was SEM. SEM has the ability to evaluate multiple relationships simultaneously, and provides more information about their strengths. However, because the linear regressions within SEM assume the input data are continuous variables rather than ordinal response scales, the results need to be evaluated with a degree of caution. (This assumption is often made with survey data; see e.g., Feys, Anseel, & Wille, 2011).

To begin this analysis, the student responses were coded numerically as +1 for positive responses (e.g., increased motivation), -1 for negative responses (e.g., decreased motivation), and 0 for neutral ones (e.g., motivation unchanged). Since the forecast grade had four categories, it was coded as +1 (80% or higher), +.33, -.33, and -1 (59% or lower), to get the same range (2) and center point (0) as the other items.

Since this study includes several different and competing hypotheses, the SEM was performed in 2 stages. The first one used a path model (Figure 1) containing all 15 hypothesized links. Table 3 shows the resulting regression coefficients, along with their standard errors, confidence intervals, and *p*-values. Table 4 provides similar information for covariances.

Table 3 indicates that the relationship between the forecast grades and the grade differences is highly statistically significant ($\beta = .341, p < .001$). Students were more likely to report a positive grade difference if their forecast grade was relatively high. The link only has slight explanatory power, however ($R^2 = 10.8\%$).

Both the forecast itself ($\beta = .458, p < .001$) and the grade difference ($\beta = .532, p < .001$) show positive influences on students' feelings, with a combined $R^2 = 29.3\%$. Students were more likely to report positive feelings if their forecast was good in absolute terms (a high mark) and/or in relative terms (a higher than expected mark). Only the grade difference shows a

significant influence on their recommendations ($\beta = 0.233, p = .004, R^2 = 4.9\%$). Students were more likely to favor forecasting if their own forecast was higher than expected.

Neither grade measure shows a significant impact on student motivation. Motivation is positively associated with recommendation (covariance = .111, $p < .001$), which in turn is positively associated with feelings (covariance = .137, $p < .001$). Students who reported positive feelings about the forecast tended to give it positive recommendations, and students who provided positive recommendations tended to report increased motivation.

Four variables display significant connections with student studying effort. Recommendation ($\beta = .211, p = .009$) and motivation ($\beta = .162, p < .023$) are both positive influences on studying, as expected. Conversely, feelings ($\beta = -.167, p = .007$) and the forecast grade itself ($\beta = -.194, p = .027$) are both negative influences. Altogether these four give $R^2 = 20.6\%$. Students were more likely to study harder if the forecasting exercise made them more worried or more motivated, if it seemed like a good idea, or if it gave them a low forecast.

The second stage of the SEM analysis used a simplified path model (Figure 2) that included only the 10 significant links ($p \leq .05$). Three metrics were used to evaluate the fit of this model to the data (see e.g., Arbuckle, 2009: Appendix C). The chi-squared statistic indicates whether there is sufficient evidence to reject a proposed model. The goodness of fit index (GFI) gives a score where larger is better, 0.95 is recommended, and 1.00 is preferred. Conversely, the root mean square error of approximation (RMSEA) indicates the “badness of fit” and gives a score where smaller is better, 0.05 is recommended, and 0.00 is preferred.

These measures show that the simplified model fits the data very well (GFI = .991, RMSEA = .000) and there is no evidence to reject it ($\chi^2 = 3.826, df = 5, p = .575$). In Figure 2,

the numbers beside single-headed arrows show the resulting standardized regression weights for the simplified model, while those beside double-headed arrows are the correlations.

To test whether the SEM results might be instructor-dependent, this model was re-analyzed using the instructor as a grouping variable. The responses were divided into two groups: those from classes taught by the author ($N_{Author} = 90$), and those from classes taught by the other instructors ($N_{Others} = 54$). The responses relating to the three other instructors were combined into a single group to avoid having very small sub-samples. The grouped model was then compared to the previous freely-estimated model. The grouped model has many small changes in individual coefficients (not shown), but all the same links remain statistically significant. A chi-squared test of the difference due to controlling for the instructor showed that it is not significant ($\chi_{diff}^2 = 15.991 - 3.826 = 12.165$, $df_{diff} = 20 - 5 = 15$, $p = .666$). This is a useful result, as it provides more confidence about the generalizability of the results.

Relationships: Comparing the Results

Both SEM and the chi-squared tests identified the same 10 pairings of survey items (see e.g., Figure 2) with statistically significant relationships. This corroboration provides some assurance that the results are not merely an artifact of one particular method and its statistical assumptions.

DISCUSSION

Support for Theoretical Hypotheses

Hypotheses 1 and 2 are well supported by the data. As expected, stronger students are more likely than weaker students to make use of grade forecasting. Similarly, students tend to over-estimate their course performance, with weaker students being more prone to this problem.

Hypotheses 7 and 8 are mostly supported as well. The students' affective and cognitive responses all impact studying behavior; however, student feelings have a negative, rather than

positive, influence on studying (Hypothesis 7a). The responses are also related to each other, except that there is no direct correlation between feelings and motivation (8c).

Hypotheses 3 and 4 regarding SRL are not well supported by the results. The difference between the forecast grade and the student's prior expectations does not display any direct impact on their studying effort (3) or motivation (4c). It does impact their feelings (4a) and recommendations (4b) about the forecasting experience.

The support for rival Hypotheses 5 and 6 is slightly stronger. The forecast grade has a negative influence on studying effort (5), and a positive influence on feelings (6a), though none on recommendations (6b) or motivation (6c).

With regards to the overall path model, 10 of the 15 hypothesized links from Figure 1 are statistically significant, and the simplified model in Figure 2 fits the data very well. The modest strengths of the coefficients may be due to latent moderating variables that are not captured by this study's simple design. For example, students having different personality types or academic strengths might react differently to the predictive feedback. If one group reacts positively while another reacts negatively, then the data as a whole could show little net effect. Future research should explore these potential influences.

Contributions and Implications

The primary goal of this study was to explore grade forecasting as a pedagogical tool. Would students be interested and influenced by this novel form of feedback, or would they ignore it? Is forecasting worth trying in future courses?

The research was successful on this dimension. About half the respondents reported trying the optional forecasting tool, and most recommended its continued use. The survey also showed modest but significant changes in students' reported behavior and emotions subsequent

to the forecast. (One student commented that after seeing his forecast, he worked harder to “beat the number”, and was very pleased when he did.)

Past studies showed that students tend to be overly optimistic about their learning, and this was true of the participants here as well. The results suggest that many but not all of these students are receptive to grade forecasting as a form of supplemental feedback.

It is reasonable to think that students who participated in this experiment also improved their self-assessment skills. As Lawrence, Goodwin, O’Connor & Önkcal (2006) noted in regards to business forecasting, “One of the key findings in the last 25 years is that feedback can be valuable because it enables the judgmental forecaster to learn.” Similarly, feedback on students’ self-assessments can help them get better at self-assessment (Nicol & Macfarlane-Dick, 2006).

For these reasons, other instructors may wish to try grade forecasting. It presumably could fit into any course where the term marks are reasonable predictors of the overall grades.

The secondary goal of this research was to test several hypotheses related to learning theory. The results in this regard were mixed. It is puzzling that, contrary to SRL theory, the difference between the expected grade and the forecast seems to have no direct impact on student motivation and studying effort. Instead, increased studying was reported by students who had experienced increased anxiety, increased motivation, or positive impressions subsequent to the forecasting experience, as well as by students who had received low absolute grade forecasts. It seems that grade forecasting does create affective and behavioral responses; but these are not as straightforward as originally thought.

It may be that student goals include not just point targets, but also ranges. For example, a student might “expect” a 75 (their point estimate), but be “satisfied” with anything between 70 and 80 (their range estimate). In that case, a forecast of 71 could “disappoint” by being lower

than expected, and yet not trigger a behavioral change because it is still within the acceptable range. This is similar to how industry uses statistical process control: each process has a target value, but the operator intervenes only if the process deviates beyond the control limits.

The third goal of this research was to verify several findings from previous studies of teaching and learning. The experiment confirmed that students in this novel context do tend to display excessive optimism about their course progress, and stronger students are more likely than weaker students to participate in forecasting.

Limitations and Future Research

One disappointment in this project was the low participation rate of weaker students. Stronger involvement from this group might have been expected, since these students could have benefited most from the extra feedback. But unfortunately that did not happen.

This low participation somewhat limits the generalizability of the study's results. The survey presumably gives a reasonable picture of forecasting's effects for the above-average and average students who participated, but probably not for the below-average students who mostly opted-out. Thus the study's conclusions may be limited to the former groups.

On the other hand, the unequal participation might make some of the conclusions more interesting. That is, although the respondents were disproportionately strong and/or involved students, they still showed excessive optimism about their grades and surprise about the forecast.

Clearly one goal for future research is to obtain greater participation by the weaker students. For the study herein, participation in the forecasting activity was optional, and the survey was administered in a class where attendance was optional. Consequently, the tendencies of weaker students to skip class (Westerman, Perez-Batres, Coffey, & Pouder, 2011) and not seek extra feedback (Butler & Winne 1995) likely limited their involvement. It also could prove

interesting to survey in more detail the students who choose not to try forecasting. Do they simply not see any value in it, or is there some other reason for their choice?

Another dimension to explore in a future is the timing of the forecast, as this involves a trade-off between accuracy and response time. The study herein provided the forecast near the end of the course, when the term marks were almost all known; this made the forecast relatively accurate. The students subsequently had about 4 weeks to prior to the exam (worth about 50% of their course grade) to respond by allocating more (or less) time and effort to studying. If instead the forecasting were to be done earlier in the semester, then students would have more opportunity to take corrective action; however, the forecasts would be based on fewer term marks and thus be less accurate. Complicating this tradeoff is the apparent tendency for students to disregard feedback given “too early” (Love, Love, Northcraft, 2010). Thus it would be interesting to find the optimal time for using grade forecasting.

It would also be beneficial for future research to include a control group of students who complete a survey but do not receive grade forecasts. This would make it easier to distinguish the extent to which the student responses are due to the forecasting experience itself, as opposed to other concurrent factors such as the approach of examinations.

ACKNOWLEDGEMENTS

Thanks are due to three anonymous reviewers for suggesting improvements in this paper.

REFERENCES

- Arbuckle JL (2009). *Amos 18 User's Guide*. Amos Development Corporation, Crawfordville.
- Beck HP, Davidson WD (2001). Establishing an early warning system: Predicting low grades in college students from survey of academic orientations scores. *Research in Higher Education*, 42 #6, 709-723

- Beyer HN (1971). Effect of students' knowledge of their predicted grade point averages on academic achievement. *Journal of Counseling Psychology*, 18 #6, 603-605.
- Butler DL, Winne PH (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65, #3, 245-281.
- Carver CS, Scheier MF (1990). Origins and functions of positive and negative affect: A control-process view. *Psychological Review*, 97, 19-35.
- Feys M, Anseel F, Wille B (2011). Improving Feedback Reports: The Role of Procedural Information and Information Specificity. *Academy of Management Learning & Education*, 10, #4, 661-681.
- Klassen K, Biktimirov E (2007). Relationship Between Student Performance and Specific Online Support Materials in an Operations Course. *Journal of the Academy of Business Education*, 8, Summer, 40-48.
- Kluger AN, Adler S, (1993). Person- versus computer-mediated feedback. *Computers in Human Behavior*, 9, #1, 1-16
- Kruger J, Dunning D (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77, 1121–1134.
- Lawrence M, Goodwin P, O'Connor M, Önkal D (2006). Judgemental forecasting: a review of progress over the last 25 years. *International Journal of Forecasting*, 22, 493–518.
- Leitner J, Leopold-Wildburger U (2011). Experiments on forecasting behavior with several sources of information – A review of the literature. *European Journal of Operational Research* 213, 459–469

- Nicol DJ, Macfarlane-Dick D (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31, #2, 199-218.
- Nowell C, Alston RM (2007). I thought I got an A! Overconfidence across the economics curriculum. *Journal of Economic Education* 38, Spring, 131-138.
- Pavlovich K, Collins E, Jones G (2009). Developing Students' Skills in Reflective Practice: Design and Assessment. *Journal of Management Education*, 33, #1, 37-58.
- Sitzmann T, Ely K, Brown KG, Bauer KN (2010). Self-Assessment of Knowledge: A Cognitive Learning or Affective Measure? *Academy of Management Learning & Education*, 9, #2, 169-191.
- Sitzmann T, Brown KG, Casper WJ, Ely K, Zimmerman RD (2008). A review and meta-analysis of the nomological network of trainee reactions. *Journal of Applied Psychology*, 93, #2, 280-295.
- Strasser SE (2003). Will This Be on the Exam? or, How to Get Students to Study More. *Decision Sciences Journal of Innovative Education*, 1, #1, 155-158
- Van Oudenhoven JP, Siero F, Veen P, Siero S (1982). Effects of positive feedback on self-evaluations and spelling achievement in a real educational setting. *European Journal of Social Psychology*, 12, 321-325.
- Westerman JW, Perez-Batres LA, Coffey BS, Poudier RW (2011). The Relationship Between Undergraduate Attendance and Performance Revisited: Alignment of Student and Instructor Goals, *Decision Sciences Journal of Innovative Education*, 9, #1, 49-67.

APPENDIX 1. EXAMPLE FORECASTING SPREADSHEET

Grade Forecasting in sections 3, 4, & 5 for Jan-Apr 2011

Use of this forecasting tool is optional and anonymous.

Type your 4 term marks into the cells below:

Test 1	20.0	out of 30
Test 2	21.0	out of 28
Assignment 1	18.0	out of 20
Assignment 2	17.0	out of 20

Based on your term marks above and the results of past courses, below is a forecast of your exam mark and course grades. It uses linear regression from past course results to estimate your future exam mark.

Exam Mark Forecast

The best point estimate of your exam mark is **69** out of 100

Overall Course Grade Forecast

The best point estimate of your course grade is **73** out of 100

Prediction intervals for the forecast:

There is a 50% probability that this will fall between	68	and	78
There is a 95% probability that this will fall between	59	and	87

Your minimum and maximum possible grades are **39** and **89**

The probability of getting an "A" (80 or more) is about **16%**

The probability of getting a "B" (70 to 79) is about **59%**

The probability of getting a "C" (60 to 69) is about **24%**

The probability of getting a "D" (50 to 59) is about **1%**

The probability of getting an "F" (49 or less) is at least **0%**

Note that "0%" means "< 0.5%"; it does not mean "impossible".

Keep in mind that this is a forecast and not a guarantee. Everyone's results will be affected by random variation from year to year and from student to student. Your own actual results will be affected by your studying between now and the exam.

APPENDIX 2. SURVEY QUESTIONS AND RESPONSES

(The values to the left of each response indicate the number of respondents and the percentage of respondents, respectively.)

Approximately what was your forecast overall course grade?

30	21%	80% or higher
67	47%	Between 70% and 79%
41	28%	Between 60% and 69%
6	4%	59% or lower

How did the forecast grade compare to your expectations?

9	6%	The forecast grade was significantly higher than I would have expected.
93	65%	The forecast grade was roughly what I would have expected.
42	29%	The forecast grade was significantly lower than I would have expected.

How did you feel after seeing your forecast grade?

44	31%	Positive: the forecast made me more confident or relaxed about my marks.
49	34%	I did not have any noticeable feelings or emotions related to the forecast.
51	35%	Negative: the forecast made me more worried or stressed about my marks.

Did the forecast affect your motivation or interest in this course?

80	55%	It increased my motivation, or I felt more interested.
54	38%	It did not have much effect on my motivation or interest.
10	7%	It decreased my motivation, or I felt less interested.

Do you recommend that this optional grade forecasting be used again in next year's course?

107	74%	Yes; offering a grade forecast is a good idea.
28	20%	Maybe; I'm not sure, or I don't think it matters much.
9	6%	No; offering a grade forecast is a bad idea.

Did your efforts in this course change after you saw the forecast?

67	46%	I am now studying more for this course than I had previously intended.
73	51%	No, I have not really changed my studying for this course.
4	3%	I am now studying less for this course than I had previously intended.

Table 1.

Breakdowns by forecast grade of students trying the forecasting tool, versus all students enrolled in the course, shown as both numbers of students and as percentages of the totals.

Forecast Grade	Students Trying		Students Enrolled	
80 to 100	30	21%	57	12%
70 to 79	67	47%	153	33%
60 to 69	41	28%	168	36%
0 to 59	6	4%	87	19%
Total	144	100%	465	100%

Table 2.

Showing *p*-values for chi-squared tests of association between pairs of survey items.

	1	2	3	4	5
1. Forecast					
2. Difference	.000				
3. Feelings	.000	.000			
4. Motivation	.149	.902	.484		
5. Recommend	.691	.009	.000	.001	
6. Studying	.000	.204	.015	.001	.017

Table 3.

Estimated regression coefficients for the 15-link path model of Figure 1.

	Regression Coefficient	Standard Error	95% Confidence Interval	<i>p</i>
Forecast → Difference	.341	.082	.180 to .501	.000
Forecast → Feelings	.458	.114	.248 to .681	.000
Forecast → Motivation	-.136	.104	-.340 to .068	.190
Forecast → Recommend	-.080	.095	-.266 to .106	.398
Forecast → Studying	-.194	.087	-.365 to -.024	.027
Difference → Feelings	.532	.110	.316 to .748	.000
Difference → Recommend	.233	.080	.076 to .390	.004
Difference → Motivation	.067	.100	-.129 to .263	.503
Difference → Studying	.015	.086	-.153 to .184	.861
Feelings → Studying	-.167	.062	-.288 to -.045	.007
Recommend → Studying	.211	.081	.052 to .370	.009
Motivation → Studying	.162	.071	.023 to .301	.023

Table 4.

Estimated covariances for the 15-link path model of Figure 1.

	Covariance	Standard Error	95% Confidence Interval	<i>p</i>
Feelings ↔ Recommend	.111	.034	.044 to .178	.000
Recommend ↔ Motivation	.137	.032	.074 to .200	.000
Motivation ↔ Feelings	.048	.036	-.023 to .119	.176

Figure 1.
Path model showing all 15 hypothesized links between survey items.

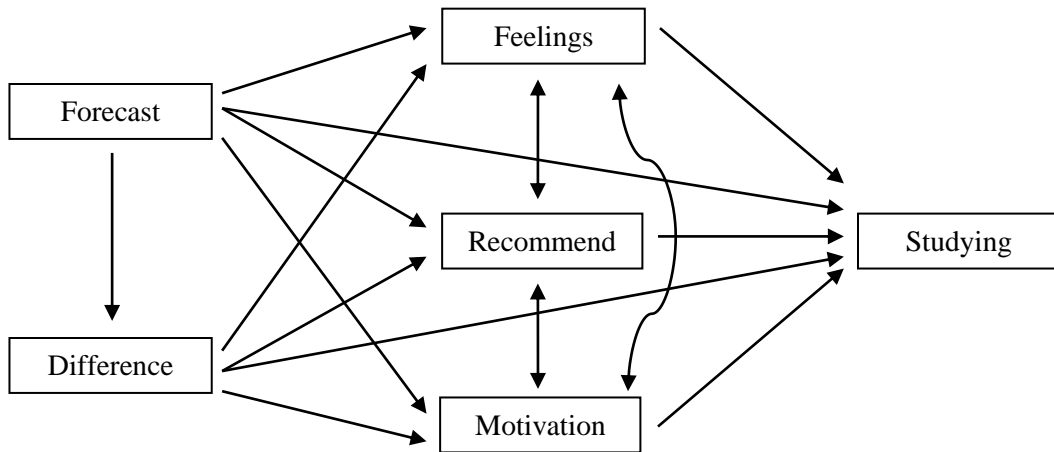


Figure 2.
Simplified path model showing statistically significant ($p \leq .05$) standardized regression weights and correlations.

