JISAE ICERE
Journal of Indonesian Student Assessment and Evaluation
The 2nd International Conference on Educational Research and Evaluation
https://doi.org/10.21009/JISAE
JISAE (Journal of Indonesian Student Assessment and Evaluation)
ISSN : P-ISSN: 2442-4919 | E-ISSN: 2597-8934
Website : http://journal.unj.ac.id/unj/index.php/jisae
Vol 9 No 2 (2023)

Formative Assessment in a Geology Classroom: *Promoting Learning by Giving Top Marks*

Daniel S. Helman[1]
[1]Faculty of Labor Relations and Trade Unions, Ton Duc Thang University, 19 Nguyen Huu Tho, Tan Phong, Quan 7, Ho Chi Minh City, Vietnam

Address for Correspondence:
helmandaniel@tdtu.edu.vn

ABSTRACT

Based on the observation that teacher expectation often drives classroom success, a project was undertaken whereby all students were given top marks in two sections of a geology classroom at the undergraduate level. To further this project, a student motivation strategy was attempted to take the place of their not having to fear for top marks. The results were sixfold: (1) the results of student work were well above average; (2) student work reflected teacher expectations; (3) student enthusiasm was well above average; (4) the instructor needed to use enthusiasm and humor to maintain discipline; (5) the instructor needed to parse the assignments so that the work would be done; (6) the administration was unhappy with the resulting data analytics. Based on this case study, one can confirm that student outcomes are highly dependent on teacher expectations and strategies; and that formative (not evaluative) assessments can be used effectively to raise the level of student learning.

Keywords: Formative Assessment In Education; STEM Teaching; Radical Pedagogy; Managing Student Expectations; Student-Centered Learning.

INTRODUCTION

There are two types of evaluation in the context of social science and pedagogy. Summative evaluation looks at the facts for a given object of study. It is called *summative* because it is meant to be the sum total of the facts. The results are a collection of details that can be added or fit into the same conceptual category. A fact-finding report, such as an environmental impact report, can be classed as a summative evaluation.

If an assessment transforms the object of its study, it is a formative assessment. Even if the assessment is rigorously and objectively applied, so long as there is a transformation, it can be classed as a formative assessment. A needs assessment, for example, is a formative evaluation, since the act of assessing student needs has transformative power over their condition in the act itself.

The above classifications make working with assessments in an educational setting particularly challenging. Expectations play an important role in the education of students. On the teacher's side: If a teacher believes that their students are highly gifted, they will become so even if the original belief was mistaken (Rosenthal & Jacobson, 1968; Jussim & Eccles, 1992). On the student's side: A student's grades impact their sense of self-worth, and can reinforce a positive or negative self-image (Jones & Grieneeks, 1970). In short, teachers who think their students are bright will teach such that their students will become brighter. Students who think that they themselves are bright will be able to do ever increasingly wonderful and creative work.

Education serves to provide opportunities for learning new knowledge and skills, but exists within an economic context that demands accountability in the form of increased test scores attributable to the educational intervention being funded. Setting up a causal study is often important for the interests of the funding body (Slavin, 2002).

Causal studies take a certain form so that study results can be compared with a control sample. At its best, a controlled sample of students would match some students to each other based on demographics, and give only one group the desired intervention. Consider the plight of the parent who finds out that their child is missing out on the opportunity to benefit from a new study! This type of set-up is unethical and should be avoided. Instead, two competing interventions are often paired up, and students can be given one or the other. Yet the intangibles involved in instruction make demonstrating that the two groups are essentially equivalent leave room for doubt. No two classrooms are the same. Each is pervaded by a spirit both of the teacher and of the group of students who happen to be present in it. As teachers, we know this to be the case, and yet we make the effort to do this type of study because it is what the funding agencies demand. We want to be able to test our programs and see the results.

Or perhaps we do not consider that the impossibility of randomization will affect our results, and we subscribe to an internal consistency because there are numerical data that have been collected, and with these we can test the results and compare what we have with a random distribution, notwithstanding limitations to the data. Many researchers take this *pro forma* approach, not considering inherent study design (Bailey & Garner, 2010).

Summative assessments are critical in some areas. One would not wish to see a medical doctor who does not know the proper anatomy terminology to be able to discuss a case. There are some things for which the public good demands a measure of. One can explicitly define what public good ought to be served, and run an assessment according to those defined measures.

Yet education fundamentally is formative, and thus the notion struck me to try an experiment in my third and fourth semesters of teaching undergraduates in a science lab class, wherein all students would get the highest marks. This paper is a description of that process and of the results that came from it.

METHODS AND MATERIALS

The experimental part of the study took place during the Fall 2008 (F08) and Spring 2009 (S09) semesters at large public university in California and involved two classes of students for F08 (F08i, F08ii) and one class for S09 (S09i) of a one-unit lab course entitled "Introduction to Geology Laboratory." The author had taught the course already for one year, with two classes in Fall 2007 (F07i, F07ii) and three classes during Spring 2008 (S08i, S08ii, S08iii). For the first semester (F07), the course was taught with a student-centered approach of designing assignments and assigning grades based on student work that was aided via both student choice and teacher support. The projects, for example, were open-ended, in that students designed their own fieldwork as a capstone activity, and the teacher coached students on how to complete the work if they wanted to expend a minimum of energy. (Of course, students could do more elaborate projects if they desired.) The course included workbook activities that were accompanied by teacher-generated notes that explained how to arrive at the correct answer.

The second semester (S08) extended this studentcentered approach by replacing some of the workbook activities with demonstrations or conceptbuilding activities that were more closely tailored to the teacher's objective for the lesson, which resulted in a wider range of perspectives with which students approached the subject, and gave additional opportunities for drawing and listening skills to students in addition to analytic and text-based challenges. Next, the second year (F08-S09) saw the replacement of the previous grading system with a secret plan to give all student work the grade of A (top marks). The syllabus states: "Formula to be announced near the end of the semester; likely, all the work is weighted equally. Extra credit or makeup assignments are (always) allowable." Student work was read and commented

on by the teacher as usual, with a mark of "A — Great Work!" or something similar added to the top.

The classes consisted of the following enrollments for F07i-ii, S08i-iii, F08i-ii and S09i, respectively: 26, 30, 31, 34, 35, 29, 29, 29. Students were undergraduates taking the course either to fulfill a laboratory science requirement—students could pick among physics, biology, chemistry or geology—or because they were interested in becoming a geology major. Most students were in their first two years of undergraduate studies, and women were more prevalent than men.

Classes met once per week for four hours in a classroom fitted out with geology equipment and supplies and laboratory benches that allowed for collaborative work in small groups if students wished. There was no lecture component to the course, but to receive credit as a laboratory science, students were required to attend a different lecture section as well. The laboratory exercises were designed with the sequence of the lecture topics in mind, so that students would have hands-on experience that would support their learning what was being taught in the four-credit lecture class.

The study herein described is a case study, and by nature is qualitative. Typically qualitative studies provide insight into novel phenomena and are used in an exploratory fashion to identify new areas to focus on in future work. Qualitative studies have the flexibility to accurately present new information in a way that is useful, and rely on the use of multiple modes of analysis to build the case for their reliability.

This study explores the following data: official student evaluations, unsolicited student opinions of the courses, online student comments, unsolicited colleague comments, unsolicited student comments, artifacts from the courses, and finally teacher reflections about the course and its outcomes.

The eight class sets of official student evaluations include a rating scale (1-5, 5 is most positive) of eight questions, and the ratings include the mean (M), standard deviation (SD) and standard error of the mean (SEM) for each of the questions both for these classes as well as for the department plus the M and SD for the university. These data are compared between the two years to identify changes in student ratings.

For the other types of data in this study, reflection and language-based analysis were employed to identify major themes present. The results were checked for obvious inconsistencies and contradictions, and assessed for their ability to provide a reasonable and useful explanation of what occurred.

RESULTS AND DISCUSSION

Quantitative Student Evaluations

The student evaluations had eight questions, listed in Table 1. The proportion of students responding for the eight classes were as follows: 12/26, 20/30, 25/31, 26/34, 26/35, 22/29, 22/29, 23/29. (N=176; n1=109 and n2=67, where n1 is normal grading and n2 is all-top-marks). Scores range from 1–5, with 5 being more positive. Scores across all questions were on average 0.29 points higher than the college mean for each question (max: 0.39; min: 0.13; SD: 0.20). Likewise, scores across all questions were on average 0.32 points higher than the department mean for each question (max: 0.43; min: 0.16; SD: 0.24). This is attributable to the author's credential training and experience in adult education, compared with most university instructors who receive no training in pedagogy.

Table 1. Student Evaluation Questions.

No.	Question	Short Version	Scale
1	Instructor provided clear and accurate information regarding course objectives, requirements and grading procedures.	Good Course Information	Strongly Agree (5)—Strongly Disagree (1)
2	The instructor's grading was consistent with stated criteria and procedures.	Fair Grading	Strongly Agree (5)—Strongly Disagree (1)
3	The instructor provided assignments/ activities that were useful for learning and understanding the subject.	Useful Course Activities	Strongly Agree (5)—Strongly Disagree (1)
4	The instructor's expectations concerning work to be done in the course were reasonable.	Not Too Much Work	Strongly Agree (5)—Strongly Disagree (1)
5	The instructor was well prepared for class.	Instructor Prepared	Strongly Agree (5)—Strongly Disagree (1)
6	The instructor was effective in pre- senting subject content and materials in the class.	Good Instruction	Strongly Agree (5)—Strongly Disagree (1)
7	The instructor was available during posted office hours for conferences about the course.	Available for Office Hours	Strongly Agree (5)—Strongly Disagree (1)
8	Rate the overall teaching effectiveness of this instructor in this course.	Overall Instructor Effectiveness	Excellent (5)—Very Poor (1)

Mean question scores that changed more than half the standard deviation between n1 and n2 were Q1 (good course information) and Q3 (useful course activities): 0.18 (SD: 0.22) and -0.16 (SD: 0.20) respectively. Notwithstanding, greater changes were seen within n1 between F07 and S08 with the adoption of more student-centered activities, on average a change of magnitude 0.29 (max mag:

0.49; min mag: 0.12; SD: 0.19). Changes were in the positive direction for all questions except Q3 (useful course activities) which was –0.12 (SD: 0.21). In short, students were slightly less likely to report that the activities were useful once they included additional non-traditional student-centered components. The change in activities had a greater impact than the change in grading on student quantitative evaluations of the course.

A correlation (r-value) array was calculated for the mean scores and standard deviations both for these course averages and for college and departmental averages, to see if the variation correlated to the semester (1, 2, 3 and 4) and/or to the grading policy (Years

1 and 2). The correlation array for the course averages and for the college and departmental averages are given as Table 2 and Table 3, respectively.

Table 2. Student Course Evaluation Mean Scores and Their Correlation Coefficients with Semester and Grading Mode.

Class	Grading	Semester	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
F07i	1	1	4.25	4.50	4.75	4.41	4.58	4.16	4.55	4.41
F07ii	1	1	4.58	4.88	4.76	4.76	4.58	4.23	4.53	4.41
S08i	1	2	4.64	4.64	4.32	4.76	4.76	4.48	4.69	4.32
S08ii	1	2	4.76	4.92	4.80	4.96	4.92	4.76	5.00	4.80
S08iii	1	2	4.84	4.96	4.80	4.92	4.96	4.80	5.00	4.92
F08i	2	3	4.68	4.68	4.40	4.72	4.77	4.54	4.73	4.50
F08ii	2	3	5.00	5.00	4.72	4.95	4.95	4.81	4.89	4.86
S09i	2	4	4.69	4.82	4.47	4.73	4.73	4.47	4.70	4.60
		Correlation with semester:	0.55	0.27	-0.49	0.31	0.41	0.46	0.31	0.32
		Correlation with grading mode:	0.42	0.16	-0.41	0.11	0.19	0.25	0.05	0.18

Table 3. Student College and Department Course Evaluation Mean Scores and Their Correlations with Semester and Academic Year.

College: Natural Sciences and Math										
Term	Year	Semester	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
F07	1	1	4.40	4.47	4.28	4.38	4.48	4.20	4.62	4.27
S08	1	2	4.45	4.50	4.33	4.43	4.55	4.27	4.65	4.34
F08	2	3	4.38	4.44	4.23	4.34	4.49	4.17	4.61	4.23
S09	2	4	4.41	4.46	4.27	4.38	4.51	4.23	4.64	4.29
		Correlation with semester:	-0.18	-0.46	-0.41	-0.32	0.13	-0.03	0.14	-0.14
		Correlation with academic year:	-0.59	-0.81	-0.77	-0.70	-0.28	-0.47	-0.32	-0.57
Department: Geological Sciences										
Term	Year	Semester	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
F07	1	1	4.39	4.43	4.19	4.34	4.47	4.25	4.60	4.26
S08	1	2	4.56	4.60	4.38	4.58	4.62	4.42	4.71	4.44
F08	2	3	4.22	4.32	4.01	4.18	4.39	4.03	4.47	4.05

S09	2	4	4.24	4.35	4.02	4.27	4.35	4.05	4.53	4.12
		Correlation with semester:	-0.65	-0.53	-0.65	-0.46	-0.64	-0.69	-0.56	-0.61
		Correlation with academic year:	-0.90	-0.83	-0.90	-0.79	-0.85	-0.93	-0.87	-0.89

There was more correlation of the changes in student response at the college and department level than there were at the course level. Changes in student evaluations are at most modestly correlated at the course level with the experiment for some questions, e.g. Q1 (0.41) and Q3 (–0.41). However, these values have greater magnitude (0.55; –0.49) when correlated by semester. The highest correlation in the course data (not shown) is a high correlation value in the standard deviation for Q3 which increases (0.66 to 0.73) when the analysis is changed from semester to grading policy.

Online Student Comments

A total of six student ratings for this course appear on the website Rate My Professor. Their provenance are as follows: F07(2), S08(3), F08(1). The ratings are shown in Table 4.

Table 4. Student Comments on Rate My Professor.

	1	2	3	4	5	6
Semester	F07	F07	S08	S08	S08	F08
Date	20/12/07	27/01/08	25/04/08	25/04/08	01/05/08	04/12/08
Rating	Good	Awesome	Average	Good	Good	Good
Overall Quality	4.0	4.5	3.0	3.5	4.0	4.0
Level of Difficulty	1.0	2.0	1.0	1.0	1.0	1.0
Comments	Geology lab 104: REALLY REALLY EASY!!!!! He doesnt lecture, he lets you leave whenever you	At first, I was a little scared of this lab class because there was no lecture or anything. But he is helpful if you ask	No Comments	This man is probably one of the strangest people you'll ever come into contact with in your life. lol. he's	Dorky kind of guy and extremely easy! He is kind of weird but really nice and helpful (he reminds me of the	helman is a great lab teacher! hes really kind and caring. his labs are easy, but they can be too long

	want, he gives you hint pages for every lab and every question, you can turn in your labs whenever you want. He is so easy so take his class.	questions. If you show that you care, he would never fail you. I got an A in the class. All you have to do is show up to get the lab assignments. He's a pretty good guy.		very passionate about everything, the class is extremely easy, and he brings you fruit to class! You have to go to class, because there's a lot of class work, which you can turn in at anytime, but you dont wanna play catch up. take this class	guys from Beauty and the geek...lol). He doesn't assign any homework and gives extremely easy exams. Just attend class everyday and do the in class assignments, that's all! I definitely recommend him!	sometimes. i totally recommend him for anyone who thinks geology is super difficult and just wants to get through it without any stress

The ratings and comments are similar throughout, with emphasis on ease and kindness, as well as professor eccentricity. The ratings are higher in the first semester before any changes had been made to either the course content or the grading system.

Unsolicited Colleague Comments

Two colleagues provided unsolicited comments about the work during the period of the course instruction. Comment 1 occurred during the first semeseter (F07). Comment 2 occurred during the third semester (F08), after the additional activies had been added to the teaching, and after the all-topmarks grading policy had been adopted. The comments are listed in Table 5.

Comment 2 is especially telling, in that this is a colleague familiar with the content of the course. The student work she saw seemed to her to be from a different, more advanced course.

Table 5. Unsolicited Comments from Colleagues.

	1	2
Title/Role	Full Professor	Teaching Associate
Comments	You are a gifted teacher. Fruit? You give them fruit before an exam? How wonderful!	What class is that for? You must be teaching an advanced class. What class is that for? The work is so excellent.

Unsolicited Student Comments

Student comments made to the teacher fell into three categories. After the transition in activity work from F07 to S08, some students considered the work to be too easy. They found the types of activities as too similar to what might be called "remedial" lessons.

After the transition in grading from S08 to F08, there was a great deal of excitement when there was a realization among students that grades were high in the class. Students viewed this very favorably. Some students did not catch on that I had a secret policy to assign the highest marks, and were active in their concern for what they needed to do to maintain their high marks! Thus the plan of both convincing students to use other (intrinsic) means of doing the learning activities and of assigning the highest marks was consistent with them actually earning these marks. A third category of student comment was the most common, and that was to express a sense during the end of the course that they had not expected to learn as much as they did, and that they have very fond memories of the class and indeed learned so very much. Many students gave personal thanks for encouraging them and for making the learning non-threatening and productive.

Course Artifacts

About 600 artifacts (student projects, tests and laboratory sheets) from the course have been scanned and saved on computer. Three were chosen at random with three numbers generated from the website Random.org.

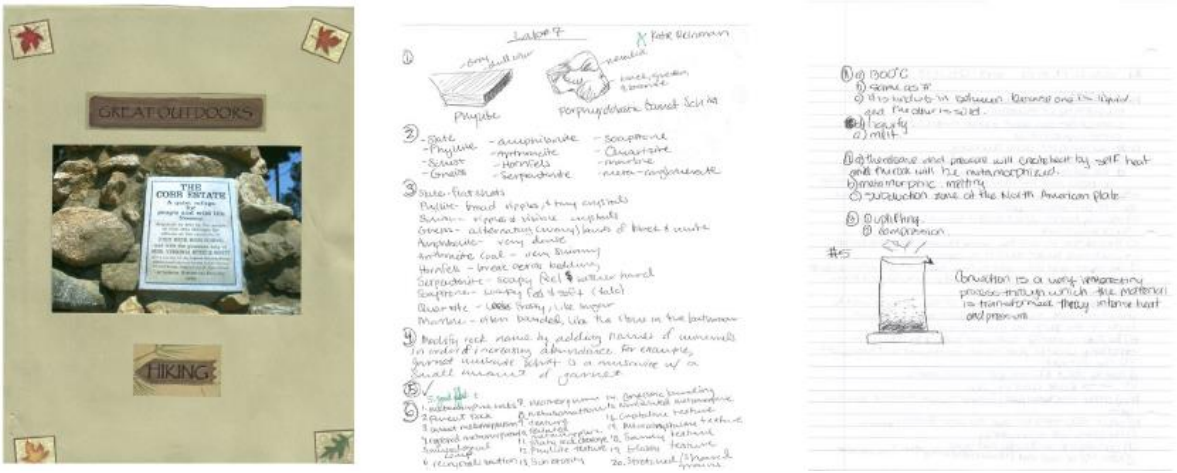

Figure 1. Student coursework.

The first image is the cover of a final project in the course. Final projects were designed by the students themselves with instructions to include at a minimum two graphs, twenty pictures, and a page of text. One student went all the way to the Grand Canyon for his project. Others did very interesting exploratory work, such as charting the bathymetry at a local pier. This type of project was undertaken throughout all four semesters of the teaching, and no fundamental difference in quality was noted between the various semesters.

The second image is the first page of a classroom laboratory assignment that was completed during the period where all assignments were given top marks. The third is the second page of a classroom laboratory assignment from a different student, also completed during the period where all assignments were given top marks.

During this time, all the work was coherent and in many cases excellent. The prevalence of drawings and other visual aids had been encouraged in the teaching , and was well-represented in the required coursework. The difference in work between S08 and F08 (i.e. the transition period from standard grading to all-top-grades) is not noticeable.

Teacher Reflections

The following is based on recollection of teaching this course, and about the mood and effort involved. It is difficult to overestimate the amount of joy and comfort that students found towards the last third of the class once they discovered what had happened. It was apparent that most students were surprised by how much they had learned, and how much they loved the subject matter. This was most obvious in the mood of the classroom. The quantitative measures discussed earlier were completely blind to this aspect of the student experience.

The chair of the deparment talked with the author about the grading about two months after the first set of grades came in. He was not pleased with the explanation that students had done all of the work required, and thus all deserved their top marks. His position was that the required work needed to be more difficult if all had succeeded. Despite this interaction, the course was taught in this fashion again without comment.

CONCLUSION

The change in classroom activities from F07 to S08 simplified the work activity for the students and allowed for better work, though was met with surprise by students. The change in grading from S08 to F08 immensely improved student enthusism for the class and especially allowed them to explore course topics without fear. The latter observation is completely missed in the official assessments for the course, and is only apparent in the social context, but there it was obvious. The overall results were sixfold: (1) the results of student work were well above average; (2) student work reflected teacher expectations; (3) student enthusiasm was well above average; (4) the instructor needed to use enthusiasm and humor to maintain discipline; (5) the instructor needed to parse the assignments so that the work would be done; (6) the administration was unhappy with the resulting data analytics once they realized that all students had received top marks.

REFERENCES

Bailey, R., & Garner, M. (2010). Is the feedback in higher education assessment worth the paper it is written on? Teachers' reflections on their practices. *Teaching in Higher Education, 15*(2), 187-198. doi:10.1080/13562511003620019

Jones, J. G., & Grieneeks, L. (1970). Measures of self- perception as predictors of scholastic achievement. *The Journal of Educational Research, 63*(5), 201-203. doi:10.1080/00220671.1970.10883983

Jussim, L., & Eccles, J. S. (1992). Teacher expectations:

II. Construction and reflection of student achievement. *Journal of Personality and Social Psychology, 63*(6), 947-961. doi:10.1037/0022-3514.63.6.947

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review, 3*(1), 16-20. doi:10.1007/BF02322211

Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research.

Educational Researcher, 31(7), 15-21. doi:10.3102/0013189X031007015