Comparing Student Achievement using Pre and Post Formative Assessment Instructional Activities

By

Shawn D. Howe

A Master's Paper Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Education – Biology

Karen Klyczek

Date

University of Wisconsin – River Falls

2012

Abstract

Improving student achievement and learning is a continual goal of education. The literature supports that formative assessment and quizzing helps with these goals. I wanted to determine if a modified formative assessment would be a better practice compared to the traditional format of formative assessment. Three sections of physical science were used in the study. Two of the three sections were the experimental group, either receiving the modified formative assessment or the traditional form, while the third section was the control. These methods rotated during the first three units of semester one so each section participated in all three methods. This process was repeated for the first three units of the second semester. At the completion of the school year, 18 students were randomly selected from each section (54 total) and their data was analyzed by determining a mean from each unit exam for each experimental method. This data was then used to run a one-way analysis of variance with a post-hoc comparison to determine if there was a statistical significance in any of the methods used. When all 54 students were used despite the differences in test means, there was no statistical significance. Further analysis showed that by removing the top nine students (based upon GPA) there was a statistical significance in the modified formative assessment. This modified form of formative assessment could potentially help educators with closing the gap between the top 25% and bottom 25% academically.

Literature Review

The term "formative assessment" was first coined in 1967 by Michael Scriven in an essay in which he compared two educational evaluations: formative and summative evaluations (Scriven, 1967). Formative assessment is defined as "a planned process in which assessment-elicited evidence of students' status is used by teachers to adjust their ongoing instructional procedures or by students to adjust their current learning tactics" (Popham, 2008). In traditional practice, formative assessments should be graded but not scored, i.e they are not counted in the student's overall grade for a class. A summative evaluation is one that is used for grading purposes and tells the teacher what the student has learned at some point in time. They are usually done at the end of a unit. Teachers do not have the ability to modify their instruction to meet the needs of the class when summative assessments are used.

Summative assessments would be the kinds of assessments traditionally associated with classroom grading. Final tests, unit exams, end-of-unit projects, and semester finals are all examples of summative assessments. Examples of formative assessments are more varied. One example might be a "think-pair-share", when students are invited to discuss their thinking about a topic in pairs or small groups, then a representative is asked to share the thinking of the larger group (Boston, 2002). Another popular form of formative assessment is the "exit card". These are index cards that students hand to the teacher, deposit in a box, or post on the door as they leave a classroom and allow students to respond to a question, solve a problem, or summarize their understanding after a particular learning experience. The teacher can then quickly read the responses, sort them into groups, and use the data to inform future instruction (Dodge, 2009). Additional kinds of formative assessment could be students writing their understanding of vocabulary before and after instruction, summarizing main ideas they've taken from lecture/readings, or interviewing students about their thinking as they solve problems

(Boston, 2002). In formative assessments, students are often frequently quizzed but the purpose is to check for understanding, rather than to add points to their overall grade.

In 1998 research done by Paul Black and Dylan William, in which they did a meta-analysis on more than 250 research studies on formative assessment, they found that the effect size of students who were exposed to the formative assessment increased from between 0.4 to 0.7 compared to students who were not given formative assessments prior to the summative assessment (effect size is the ratio of the average improvement in test scores in the innovation to the range of scores of control groups of students on the same tests). This increase is equivalent to going from the 50^{th} percentile to the 65^{th} percentile (Greenstein, 2010). Another study, which compared weekly point value quizzes and test performance in a statistics class over a four year period, found that frequent quizzing with point value had a favorable and significant effect on student performance compared to a control group that was not exposed to weekly quizzes. In fact, correlation between quizzes and student test performance was 0.91 (near linear relationship). Further, the study found that there was a positive relationship between weekly quizzes and test scores and weekly quizzes and grade earned in the course. Additionally, there was a positive relationship in terms to higher test performance for individuals who were not as strong academically (Kamuche, 2005).

Research done by Henry L. Roediger III and Jeffrey D. Karpicke supports the aforementioned research, indicating that quizzing on previously learned material results in students better understanding the material compared to just studying alone. In their research, two groups were given the same material to learn. One group received three immediate free-recall tests and the other group restudied the material the same number of times as the group receiving the free-recall tests. Their results showed that the group receiving the free-recall test remembered around 60 percent of the material where the group given only study time remembered around 40 percent (Roediger & Finn,

2010). These results showed a 20 percent improvement in retention by being exposed to quizzing prior to the summative assessment.

Formative assessment is not without limitations, however. This type of assessment generally has little or no point value assigned. It should foster the idea of "wanting to learn", but sometimes students may need more motivation. Research done by Edward J. Palmer and Peter G. Devitt took fourth year medical students and divided them into four groups. Two of the four groups received formative assessment material online with feedback, one group received written formative assessment material, and the last was the control. Their results showed that student performance was not affected by any method used in the study despite the availability and encouragement to use these formative tools. The researchers further concluded that in order for formative assessment to work students may need to be motivated to use such tools by implementing some form of summative assessment (Palmer & Devitt, 2008).

Purpose

The purpose of my project was to compare two different types of formative assessments. Specifically, a traditional form of formative assessment in which a student's assessment would be graded but not scored and a non-traditional form in which the assessment would be graded and scored but the point value would be minimal. According to research, formative assessments do help student achievement assuming they are motivated to learn and use the feedback from the assessment. I want to find out if student achievement on test scores can increase with a modified form of formative assessment by placing some point value on it. My initial assumption is that by placing a point value on the assessment it will help increase student motivation to plan and prepare for this assessment which in turn will increase student test scores.

Methods

Three sections of a first year science course per semester (physical science) were used in the study. Each section consisted mainly of 9th grade level students with some 10th grade level. 61 students took part in the 1st semester (Chemistry) and 67 students in the 2nd semester (Physics). However, some 9th grade students did not take part in both semesters based upon the switch that takes place from one semester to the next. Also, based upon the number of semesters each 10th grade student needed to recover, some participated in one or both. Data were recorded in the first three teaching units of each semester. Each section for the duration of a unit would either be the experimental group, receiving either the homework quiz or exit card, or control group that received no formative assessment (see Table 1 for distribution).

Each week, the experimental group would be given the formative assessment two times. For those receiving the homework quiz, scores were recorded and for individuals receiving the exit card, the exit card was scored but not graded. Homework quizzes were given at the start of class on learning objectives and learning goals from previous lectures. After the homework quizzes were complete, they were corrected by the students and turned in to the teacher. Exit cards were completed at the end of the lecture and students were asked both open ended questions and specific questions to determine their understanding on specific learning objectives. Upon completion, students turned in their exit cards to the teacher for review after class. After the data was analyzed for each experimental method, teaching practices were modified to fit the need of each individual class based upon formative assessment practice. See Appendix A for example homework quiz questions and exit card questions. At the completion of each semester, a survey was given to each class regarding each experimental method. The survey can be found in Appendix B.

Data was analyzed by determining the test mean for each experimental method. For example, table 1 shows that 3rd hour for unit 1, 4th hour for unit 2, and 7th hour for unit 3 were given homework quizzes. An experimental mean was then calculated using the test scores from that specific method (see table 2). This was repeated for exit cards and the control for both semesters. With the calculated means for each method, a one-way analysis of variance (one-way ANOVA) with post-hoc comparison was used to determine if there was any statistical significance among methods used. In order to do this type of comparison, sample size needs to be consistent. 18 students were randomly selected from each section for a total of 54 scores for each experimental method used. If data was significant ($p < 0.05$) a Tukey analysis was done. Statistical software *Statistix 8.0* was used for calculations.

Table 1. Formative Assessment Distribution

Fall Semester			
Unit	*3rd Hour*	*4th Hour*	*7th Hour*
1	Homework Quiz	Exit Card	Control
2	Control	Homework Quiz	Exit Card
3	Exit Card	Control	Homework Quiz

Spring Semester			
Unit	*2nd Hour*	*4th Hour*	*8th Hour*
1	Homework Quiz	Exit Card	Control
2	Control	Homework Quiz	Exit Card
3	Exit Card	Control	Homework Quiz

Results and Discussion

My results indicate that in semester 1 there was no statistical significance between any of the three methods used. The average of test scores for students exposed to homework quizzes was higher than that for either the exit card or control groups (Table 2 and Figure 1). However, an analysis of variance (ANOVA) indicated that these differences were not statistically significant (P=0.20041, Table 3). For any type of statistical significance this value needs to be less than 0.05. A Tukey test was run to

confirm this result, which showed that all three means were in the same statistical group and there was no significant difference between them (Table 6). Results from semester 2 showed no significant differences between the average test scores in the three groups (Table 4). Student raw test score data for semesters 1 and 2 can be found in Appendix C.

Table 2. Experimental Method Test Mean for Semester 1

Variable	*Mean*
Homework Quiz	71.907
Exit Card	65.852
Control	67.889

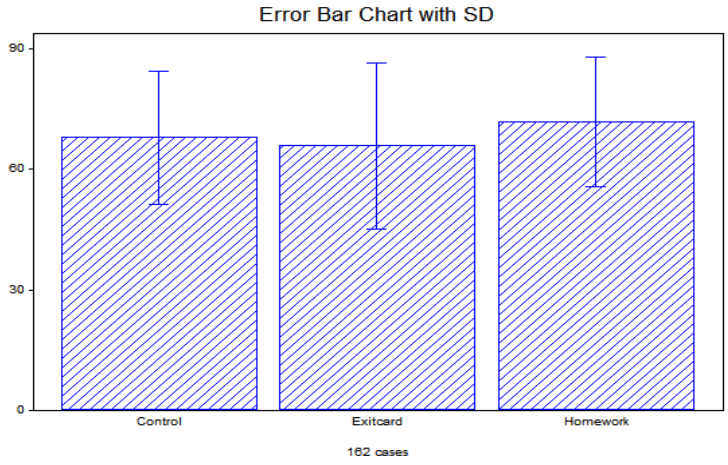
Table 3. F-Stat and P-Value for One-way ANOVA Semester 1 (all 54 participants)

F-Stat	*P-Value*
1.61	0.2041

Table 4. Experimental Test Mean for Semester 2

Variable	Mean
Homework Quiz	73.944
Exit Card	73.926
Control	72.741

Figure 1. Error bar chart with Standard Deviation for each experimental method for Semester 1 (all 54 participants).



The initial question posed by this project was to determine if a point value formative assessment was a better practice than a no point value type of formative assessment in terms of increasing student test scores. My results indicate that in semester 1 and semester 2 there was no statistical significance in either experimental method. The calculated p-value for semester 1 was equal to 0.2041 which is greater than 0.05 which would indicate any type of statistical significance. However, when looking at the mean for each method used for semester 1, there was a slight advantage for individuals when they were exposed to homework quizzes compared to exit cards and the control (see table 2). In fact, when individuals were exposed to exit cards, the control group outperformed them by approximately 2 percentage points. The results for semester 2 also indicate no statistical significance.

These results surprised me due to the content that was being covered. Semester 2 we cover general physics which means that there is math skills needed. Since formative assessment is designed for teachers to modify instruction, being able to monitor student performance in regards to their math skills I would have guessed that during these math based units, I would have been able to recognize problems in addition to give students more practice on solving various physic problems. With this particular group of students, this was not the case.

In an extension of my analysis, I wanted to see if there was any correlation to performance when compared to student GPA. Assuming the top 25% of the class will be successful despite the method used I decided to run 2 more tests. In the first test I removed the individuals with the top 6 grade point averages. There were greater differences between test means (Table 5 and Figure 2), but the differences were still not significant (Table 5). I then removed the next 3 top grade point averages (9 total). Mean differences were greater (Table 6) than the previous comparison. A Tukey comparison was done with these results and showed there was statistical significance among experimental practices

(Table 8). The question that can be asked now, if we are trying to bridge the gap between the bottom 25% and the top 25%, could the use of homework quizzes help do that?

Table 5. Experimental Test Mean for Semester 1 with Top Six GPA's Removed

Variable	Mean
Homework Quiz	70.617
Exit Card	65.362
Control	64.426

Table 6. Experimental Test Mean for Semester 1 with Top Nine GPA's Removed

Variable	Mean
Homework Quiz	69.643
Exit Card	63.143
Control	60.69

Table 7. Tukey HSD All-Pairwise Comparison Test for all 54 participants in Semester 1

Variable	Homogeneous Groups
Homework Quiz	A
Control	A
Exit Card	A

Table 8. Tukey HSD All-Pairwise Comparison Test for 9 top GPA's removed in Semester 1

Variable	Homogeneous Groups
Homework Quiz	A
Control	AB
Exit Card	B

Figure 2. Error Bar Chart with Standard Deviation for top Six GPA's removed for Semester 1

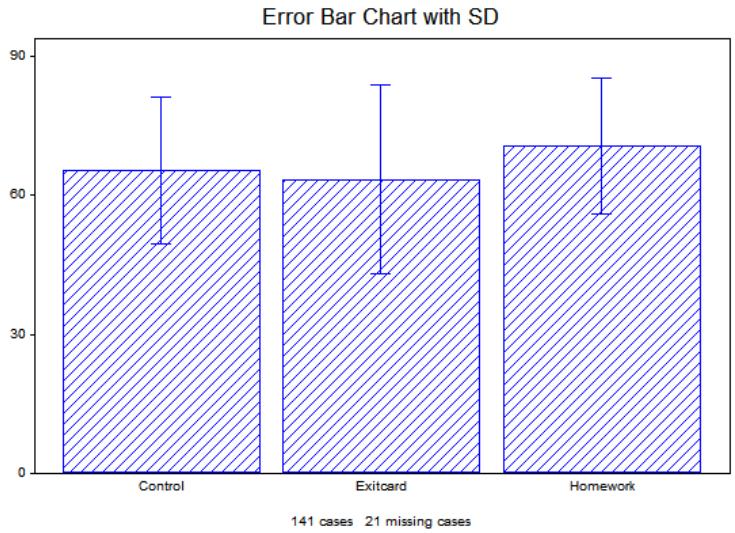
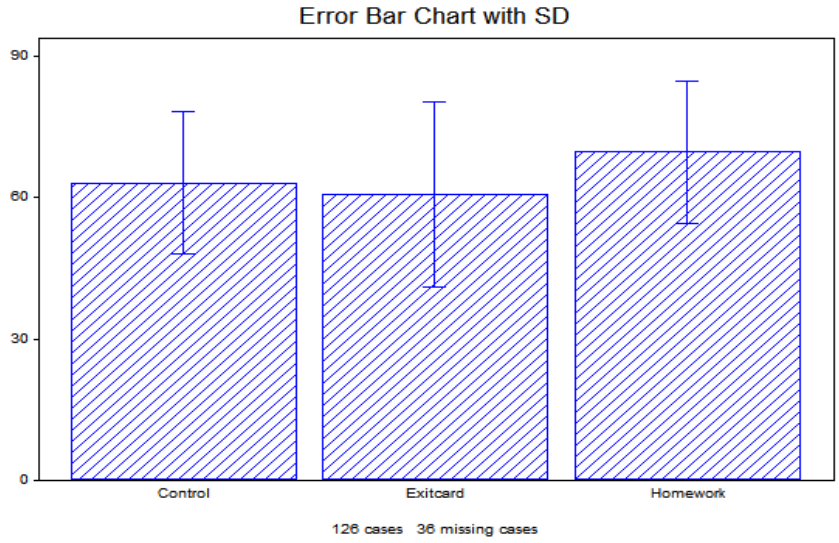


Figure 3. Error Bar Chart with Standard Deviation for top Nine GPA's removed for Semester 1



Some unexpected observations were that I would have guessed that my initial question would have been correct in that point value formative assessments would have been statistically more beneficially than no point formative assessment. The literature indicates that formative assessments do increase student achievement as does weekly quizzing. Another area of interest was the results obtained from the use of exit cards. In the comparison of means, the control group outperformed individuals when exposed to exit cards. The literature indicates that this should not be the case but for this particular group of students it was. More data would need to be collected to retest this comparison. In a positive observation, I was able to witness students on homework quiz days looking over their notes prior to the start of class. This is good practice for students to be successful.

Some factors to consider are the complexity of dealing with humans. There are many factors that I don't have control of- such as socioeconomic status, nutrition, sleep habits, substance abuse, extra-curricular activities, family obligations, academic abilities, attendance, etc. All of these can play a role in student success and achievement and when looking at the raw data, there were many outliers in

terms of individuals who would do very well or very poorly on one test and have a plus or minus difference of 20 on the next. Is this due to how well I prepared them or a lack of preparation on their part? Was it the formative assessment practice that made for these sometimes large differences or was it the particular material being covered?

As I mentioned earlier in the paper, I conducted an informal survey regarding the exit cards, homework quiz, and control. In each section, over 70% preferred the homework quizzes compared to the exit cards. The most common response was that point value helps and homework quizzes were more specific in terms of the learning goals and learning objectives. They felt that with exit cards there was a little too much freedom in terms of what they could write and not having any point value also helped decrease its importance despite the number of times I would model for them how to correctly write an exit card. One improvement I would do regarding my methodology would be to continue to collect data from year to the next. I would like to continue to collect the data as I previously did but to gather more of it over the years. The more data that can be gathered regarding this comparative study, the better you can understand best, good, and poor practices. Another possibility would be to do a comparative study of a homework quiz with point value and the exact quiz with no point value. Would students see benefits from both methods or would the results be similar to what I obtained? Another question that should be given more thought was the extended analysis of improving the bottom 25% academically and closing the gap to the top 25%.

Appendix A

Sample Exit Card questions

1. Give one example of a heterogeneous and homogenous mixture.
2. What were two questions you had from the lesson today? (Used many times)
3. List three things you learned in the lesson today. (Used many times)
4. What were two things you learned from the density lab? Be specific.
5. What is one question you still have about density?
6. Why does ice float? Provide evidence.
7. What does the percent difference tell us?
8. List one example of a physical and chemical change. You can't use examples that were used in class.
9. What is the difference between a chemical and physical change?
10. After completing the lab, what is the relationship between pressure and volume?
11. What could potentially happen if you brought an unopened bag of chips to a much higher elevation? Explain your answer.
12. What were two things that you have learned so far regarding friction and/or Newton's Laws?
13. What is the buoyant force?
14. What does the slope of a velocity vs. time graph tell us? Provide evidence in your explanation.
15. What does the slope of a distance vs. time graph tell us? Provide evidence in your explanation.

Sample homework quiz questions

1. What is the difference between a heterogeneous and homogenous mixture? Provide an example of each.
2. What is density?
3. Mathematically, how is the density of an object determined?
4. Why does ice float? Provide evidence.
5. Calculate the percent difference of the following results. What does this value tell us?
6. What are two characteristics of a physical change?
7. What are two characteristics of a chemical change?
8. If the volume of a container increases, what happens to the pressure? All other variables remain constant.
9. If temperature decreases, what happens to the pressure of the container? All other variables remain constant.
10. What is friction?
11. Calculate the slope from the following points on the distance vs. time graph. What does this value tell us?
12. Calculate the slope from the following points on the velocity vs. time graph. What does this value tell us?

Appendix B

Student Survey on experimental methods after completion of study

1. Of the two methods used, homework quiz and exit card, which method did your prefer?
2. Specifically, why did you choose the method that you did?

Appendix C

Table 9. Student raw test scores per method for semester 1

Student	Homework Quiz	Exit Card	Control	Student	Homework Quiz	Exit Card	Control
1	41	71	38	28	75	67	86
2	71	60	55	29	79	87	76
3	44	69	60	30	53	41	73
4	56	54	55	31	88	79	88
5	79	63	63	32	90	85	90
6	91	88	81	33	100	100	92
7	81	92	69	34	63	81	77
8	79	98	73	35	90	92	97
9	64	67	60	36	73	64	87
10	89	99	89	37	85	47	81
11	73	85	53	38	68	55	56
12	71	74	58	39	74	60	66
13	66	41	47	40	60	51	64
14	96	92	90	41	47	26	35
15	35	0	32	42	70	49	46
16	48	41	43	43	79	60	58
17	79	80	71	44	91	81	91
18	77	88	83	45	74	67	69
19	98	87	93	46	77	35	56
20	57	50	60	47	50	34	67
21	83	73	73	48	71	77	83
22	75	66	75	49	87	88	92
23	65	75	58	50	77	57	56
24	51	64	62	51	45	43	58
25	65	64	45	52	87	53	75
26	75	67	68	53	84	43	71
27	94	77	76	54	43	49	46

Table 10. Student raw test scores per method for semester 2

Student	Homework Quiz	Exit Card	Control	Student	Homework Quiz	Exit Card	Control
1	56	47	64	28	100	97	97
2	60	37	68	29	95	93	88
3	61	54	72	30	91	86	94
4	93	88	68	31	68	66	48
5	56	56	61	32	91	88	94
6	93	62	76	33	98	97	99
7	70	69	64	34	100	100	92
8	9	33	32	35	44	53	60
9	81	97	100	36	95	74	93
10	86	92	85	37	58	61	30
11	66	63	66	38	100	95	97
12	70	94	85	39	88	87	66
13	53	63	70	40	75	76	73
14	73	79	76	41	80	78	58
15	88	82	91	42	69	70	83
16	70	88	85	43	63	48	57
17	52	79	87	44	65	70	74
18	74	80	95	45	94	95	89
19	78	60	48	46	62	76	76
20	53	55	50	47	71	81	50
21	98	93	94	48	60	27	48
22	70	71	67	49	63	81	64
23	72	76	67	50	92	95	91
24	58	58	67	51	37	40	0
25	72	56	65	52	80	85	88
26	98	91	82	53	88	94	71
27	100	97	100	54	56	59	63

References

Boston, C. (2002). The concept of formative assessment. *Practical Assessment, Research & Evaluation, 8(9)*. Retrieved from http://PAREonline.net/getvn.asp?v=8&n=9

Dodge, J. (2009). *25 quick formative assessments for a differentiated classroom*. New York, NY: Scholastic.

Greenstein, L. (2010). What teachers really need to know about formative assessment. Alexandria, VA: Association for Supervision and Curriculum Development (ASCD).

Kamuche, F. U. (2005). Do weekly quizzes improve student performance? The Free Library. Retrieved from http://www.thefreelibrary.com/Do+weekly+quizzes+improve+student+performance%3F-a0138703686.

Palmer, E. J. & Devitt P. G. (2008). Limitations of student-driven formative assessment in a clinical clerkship. *BioMed Central.* Retrieved from *http://www.biomedcentral.com/1472-6920/8/29*

Popham, W.J. (2008). *Transformative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Roediger, H.L. & Finn, B. (2010). The pluses of getting it wrong. *Scientific American Mind, 21(1)*, doi: 10.1038/scientificamericanmind0310-38

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagné & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (Vol. 1, pp. 39-83). Chicago, IL: Rand McNally.