Final Project

Assessing the Effectiveness of Formative Evaluation in Teaching Mathematics Online

Tim Xeriland

Michigan State University

Introduction

Formative assessment as a topic of research interest can be followed back to Scriven's (1967) description of *formative evaluation* in education (Allal & Lopez, 2005). Bloom (1969) transferred the term formative from evaluation to assessment (Dunn & Mulvenon, 2009) which, according to some researchers, was the beginning of the complications in the field. The widespread confusion seems to come from a lack of agreed upon definitions. Careful examination of the formative assessment demonstrates a vagueness of operational definitions.

A look at the definition of *formative* shows a connection to the idea of forming or shaping something and achieving a beneficial result. The main idea of *formative assessment* parallels this definition which is to "shape and improve the students' competence" (Sadler, 1989, p. 120). The idea is to use assessment not to grade, but to provide feedback to the student that gives specific guidance and requires the student to utilize this guidance (Black & Wiliam, 1998). A good general definition that encompasses most of the ways formative assessment is used in the literature, (Chappuis & Stiggins, 2002; Waterman, 2010) is "assessment for learning" (Dunn & Mulvenon, 2009, p. 3).

Beneath formative assessment lie cognitive and motivational theories. The understanding of these theories is crucial to explaining how formative assessment works and why. This understanding is also important because the theoretical basis for formative assessment ties together separate elements of effective practice and helps to clarify them and see how they fit together.

Looking at the cognitive perspective the constructivist believes that learners make new information when it is incorporated in mental constructs or schemas (Smith, diSessa, &

Roschelle, 1993). With this approach a teacher's job is to help students' organize the way in which their schemas are connected (Donovan & Bransford, 2005). This is accomplished by encouraging the student to interact with where they are currently and where they need to be. This can come from teachers, peers, or even carefully constructed diagnostic feedback (Reveles, Kelly, & Durán, 2007). This interaction is most successful when it occurs close to the boundary of what a student already knows and does not yet know, called the zone of proximal development (ZPD) (Vygotsky, 1978). For Vygotsky (1978) learning is a social process that requires interacting with others.

Vygotsky's theory distinguishes two levels of development. The first is the current level of the individual. The second is the level of potential development. This potential development is the level which the student is capable of achieving with teacher and peer interaction (Heritage, 2010). Guiding the student in the process is most successful after a scaffolding has been built for the student (Belland et al., 2008). For example a teacher may ask "leading or probing questions to elaborate the knowledge the learner already possesses, or providing feedback that assists the learner to take steps to move forward through the ZPD" (Heritage, 2010, p.8). The scaffolding can be reduced as the learner becomes more competent in the subject matter until the student is able to function independently (Heritage, 2010; Tharp & Gallimore, 1988).

Formative assessment enables teachers and students to consistently work in the zone of proximal development (Heritage, 2010). By its nature formative assessment involves an interaction or a dialogue between teacher and student and it is this interaction that helps to build what Vygotsky (1978) calls "maturing factors" (p. 86). Careful utilization of formative assessment allows a teacher to provide useful experiences to support learning and to determine what is within reach of their students (Heritage, 2010). In short, formative assessment helps students assimilate new information into their schemas and thereby follows the theory laid out by the constructivist approach.

For Sadler (1989) feedback is the crucial factor to aid learning. It is usually defined by giving information about how something is being done. Sadler (1989) instead prefers a systems perspective of feedback and uses Ramaprasad (1983) definition “information about the gap between the actual level and the reference level of a system parameter that is used to alter the gap in some way” (p. 4). With this approach feedback loops are important and formative assessment is the key to closing the gap.

Practicing in a supportive environment where a teacher understand the skills necessary for improvement and how to correct poor performance is key in creating these feedback loops. For Sadler’s model, the feedback is used both by teachers to make programmatic decisions and also by students to monitor their strengths and weaknesses. The key is feedback “*when it is used to alter the gap*” and is useless if the “information is simply recorded” (Sadler, 1989, p. 121).

Feedback in the form of praise is problematic; Effective feedback focuses on the task and provides the student with suggestions, hints, or cues (Kluger & DeNisi, 1996). For the feedback to be successful it must be prompt and accurate. The closer the feedback comes to the assessment the more effective it is for student achievement (Waterman, 2010).

Further evidence points to the role effective feedback plays in the learning process. In a review of 196 studies describing nearly 7,000 effects it was reported that feedback had an average effect size of 0.79 (Hattie & Timperley, 2007; Heritage, 2010). If realized, this size effect would be one of the largest in educational interventions and would have a profound impact on a student's success. Sadler, however, was concerned with the common but puzzling observation that there may be no improvement even when feedback is provided by teachers.

Attribution theory is a good place to start when looking at what motivates students academically. It focuses on how people attribute the causes of events and how these judgments influence internal perceptions. It is tied with the concept of self-efficacy, which are students' beliefs in their own competence. The students' self-efficacy will play a role in how they will interpret success or failure.

The self-efficacy in mathematics can be defined as the internal perceptions of the student's potential for their own success (Wolters & Rosenthal, 2000). It has been shown that students with higher levels of self-efficacy set higher goals, apply more effort, and persist longer on difficult tasks (Tanner & Jones, 2003). Students with this high self-efficacy do not believe that mathematical ability is fixed and they attribute their success or failure to internal factors (Black, 1998). Black (1998) worries that when a student repeatedly gets low grades it may cause "a shared belief between them and their teacher that they are just not clever enough" (p.43). Thus, the unintended result of summative assessment can be low self-efficacy (Tanner & Jones, 2003). This can create a vicious cycle where a student fails, lowering self-efficacy resulting in more failure.

A study by Butler (1988) demonstrated that high grades boosted student interest. As the grades dropped, interest plummeted. Another study by Butler (1987) broke 200 students into four groups. Individuals from each group were given one of four kinds of feedback on a given lesson: comments, grades, praise, or no feedback at all. Only students who received comments showed improvement on the second lesson over the first.

The experimental hypothesis can be summarized as follows:

1) Post-test interest and performance scores will be highest for the group that received comments only.

2) Post-test interest and performance scores will have a greater affect on low performing students in comments only group.

3) Post-test transfer of learning will be highest for group that received comments only.

4) Post-test performance scores will show a larger gain compared to other groups for low performing students of comments only group than for high performing students of comments only group.

Method

Participants

The participants were composed of 90 students drawn from the Dallas County Community College District (DCCCD) which is the largest undergraduate institution in the state of Texas. Recruiting students was done through multiple methods drawing students from various areas of the college. The sample comprised 51 females and 39 males with a mean GPA of 3.37. A matched pairs approach was used to randomly assign each of the participants to one of the three experimental groups. These groups were 1) students who received grades only, 2) students who received comments only, and 3) students who received grades and comments. The sample had a bimodal distribution and was further broken into high performers and low performers in each of the three experimental conditions.

Measurements

Three dependent variables were examined in this study: performance, interest, and transfer of learning. The performance variable was measured with scores on the final multiple-choice exam. The interest variable was measured with a questionnaire and by tracking mouse clicks. The transfer of learning variable was measure with a test item on the final multiple choice exam. All these measures are tracked and recorded through the Blackboard Learning Management System.

Procedures

The experiment consisted of three sessions conducted entirely online. The first session involved watching a media presentation on triangles. Followed by a quiz where all groups were required to show their work. Students were given instructions on using keyboard symbols to enter steps to solving problems digitally. Those in the group that received grades only were given a grade and a password in order to proceed the next module. Those that received comments only were given feedback on incorrect responses that they were then required to go in and address before moving on (a maximum of two rounds of this would take place before allowing the student to move on). Those in the grades and comments category received a grade, comments, and the password to the next module.

The second session ran identical to the first except it involved content on circles. After the second quiz on circles was complete the evaluator sent the password to access the multiple-choice final. At this point all evaluation was automated. In other words, students automatically received a grade for the final and took an interest questionnaire.

Analysis

For this study an experiment was run that produced 20 variables. The four main variables of interest were: final quiz score (performance), question number eight from final quiz (transfer of learning), questionnaire total (interest), and number of clicks in extra section content area (interest). Because the subjects were broken into three groups an ANOVA was conducted to see if there was a significant difference between the four variables mentioned. In addition, the bimodal distribution made it convenient to break each group into high and low performers. To look at this closely, a two-way ANOVA was conducted where the aim was to look for significance with interaction and main effects.

Results

Performance measure

A two-way ANOVA was conducted that examined the effect of feedback condition and performance on geometry score. The dependent variable, geometry score, was normally distributed for the groups formed by the combination of the feedback and performance as assessed by the Shapiro-Wilk test. There was homogeneity of variance between groups as assessed by Levene's test for equality of error variances. There was a significant interaction between the effects of feedback condition and performance on geometry scores, $F(2,84) = 5.917$, $P = .004$. Simple main effects analysis showed the grades-only condition had significantly lower geometry scores when performance level was low ($P = .000364$). In addition, the grades and comments group also had significantly lower geometry scores when performance level was low ($P = .00002$). For the comments-only group there was no statistical difference between low and high performers ($P = .971$).

Transfer of learning

A two-way ANOVA was conducted that examined the effect of feedback condition and performance transfer of learning. The dependent variable was transfer of learning. This test showed no significant interaction and was rerun with just the main effects. The groups had a *P* = .046, but the Levene's test was significant meaning the assumption of homogeneity of variance was violated. In this case the P-value was close to .05 and the Levene's test was extremely significant so careful testing needed to be done to avoid a Type I error. Because the grades-only group and the grades and comments group had identical mean scores (.63) for this dependent variable an Independent Samples T-Test was run to see if the difference between the two means was significant. When equal variance was not assumed the T-test showed the comments-only group did significantly better on the transfer of learning geometry question (*P* = .038).

Interest

A two-way ANOVA was conducted that examined the effect of feedback condition and performance on questionnaire score. The dependent variable, questionnaire score, was normally distributed for the groups formed by the combination of the feedback and performance as assessed by the Shapiro-Wilk test. There was homogeneity of variance between groups as assessed by Levene's test for equality of error variances. There was no significant interaction between the effects of feedback condition and performance on questionnaire score, $F(2,84) = 0.48$, $P = .953$. Simple main effects analysis showed the comments-only group had significantly higher questionnaire scores compared to the grades-only group (*P* = .012) and the grades and comments group (*P* = .013). There was no significance between the grades-only and the grades and comments groups (*P* = .968).

Discussion

The results confirm the importance of utilizing formative feedback in instruction especially for low performing students. As hypothesized, those who received this feedback did significantly better on the final and the results were particularly pronounced for low achieving students. The caveat is that this requires quite a bit more effort. During the study I began to dread each time a subject from the comments-only group completed a module because I knew the amount of work that was required. While the grades only group could be done in under a minute. Working closely with the comments-only group, it was clear that many students completely misinterpreted or misunderstood the content that was presented. It was the job of the evaluator to give feedback that would address this misinterpretation. If this is to be expected of teachers then professional development would be required and an understanding of the increased work load.

When talking about formative assessment, I often hear concerns about students not liking it. The results from the questionnaire suggest this is not the case. The comments-only group gave significantly higher scores on the questionnaire measuring interest. The other variable to measure interest is noteworthy. In this case it was click on the extra content area. Although there was no statically significant difference between the groups looking at the estimated marginal means plot we can see it was reversed from all the variables that were plotted. One potential confounding variable was that the comments-only group was moved to a Google Document after taking a quiz where it was more convenient to give feedback and have the student then make changes. The fact the student was away from the site could be the reason. However, I would hypothesize that the reason was the students in the other groups were intentionally seeking out interaction that the comments-only group was already receiving. Further studies would be needed to determine if this is the case.

Another hypothesized result was that formative feedback would result in transfer of learning. The last question on the final was about finding the volume of a cone. This was not discussed in any of the content and there was a sudden jump from two-dimensional objects to three-dimensional. The belief was that students in the comments-only group would do better because they had learned the material on a deeper level. In other words, if you really understand the concept of area it is easier to move to the concept of volume. Because this study was held over such a short duration it was questionable whether an effect would be seen. However, the comments-only group did significantly better at answering this question. Since we are looking for more than tricks to improve test scores this is an extremely important finding. Although this was relegated to a single question, so further testing is needed.

The results of this study seem to make it worthwhile for instructors to consider incorporating this approach into online teaching. However, many of the learning management systems, including BlackBoard, do not offer the capabilities to handle this approach effectively. For example, while there is a place to leave comments when grading a test it is not a useful option for the student to act on those comments. In other words, if a student reads the comments and takes the test again it wipes out everything and starts fresh. Unless the student prints the comments they cannot go back and access them. In addition, the test is built to be graded. The test is flagged as if nothing has been done until an actual grade is assigned making it inconvenient for both the teacher and student. Based on the results of this study it is recommend to make modification to learning management systems.

References

Allal, L., & Lopez, L. M. (2005). Formative assessment of learning: A review of publications in French. In OECD (Ed.), *Formative assessment: Improving learning in secondary classrooms* (pp. 241-264). France: OECD Publishing.

Belland, Brian., Glazewski, Krista D., and Richardson, Jennifer C. (2008). A scaffolding framework to support the construction of evidence-based arguments among middle school students. *Education Tech Research Dev*., 56, 401-422.

Black, P., & Wiliam, D. (1998). Inside the black box. *Phi Delta Kappan*, 80(2), 139-148. Bloom, B. S. (1969). *Some theoretical issues relating to education evaluation.* In: TylerRW,(ed.) Educational Evaluation: New Role, New Means. Chicago, IL: University of Chicago Press.

Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology*, 58, 1-14.

Butler, R. (1987). Task-involving and ego-involving properties of evaluation: effects of different feedback conditions on motivational perceptions, interest and performance. *Journal of Educational Psychology,* 79, 474-482.

Chappuis, S., & Stiggins, R. J. (2002). Classroom assessment for learning. *Educational Leadership*, 60(1), 40-43.

Donovan, M. S., & Bransford, J. D. (2005). Introduction. In M. S. Donovan & J. D. Brandsford (Eds.), *How students learn: History, mathematics, and science in the classroom* (pp.1 28).Washington, DC: National Academic Press.

Dunn, K., & Mulvenon, S. (2009). A critical review of research on formative assessment: The limited scientific evidence of the impact of formative assessment in education. *Practical Assessment, Research & Evaluation*, 14(7), 1-11.

Hattie, J., & Timperely, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81-112.

Heritage, (2010). *Formative assessment and next-generation assessment systems: Are we losing an opportunity?* Washington, DC: Council of Chief State School Officers. Retrieved July 30, 2011 from http://www.ccsso.org/Resources/Publications/Formative_Assessment_and_Next Generation_Assessment_Systems.html

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119, 254-284.

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, 28, 4-13

Reveles, J. M., Kelly, G. J., & Durán, R. P. (2007). A sociocultural perspective on mediated activity in third grade science. *Cultural Studies of Science Education*, 1, 467-495.

Sadler, D. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, and M. Scriven (Eds.), *Perspectives of curriculum evaluation*, Volume I (pp. 39-83). Chicago, IL: Rand McNally.

Smith, J., Disessa, A., & Roschelle, J. (1993). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences*, 3(2), 115-163.

Tanner, H. & Jones, S. (2003). Self-efficacy in mathematics and students' use of self-regulated learning strategies during assessment events. Proceedings from 27th International Group for the Psychology of Mathematics Education. PME: HI.

Tharp, R. G., & Gallimore, R. (1988). *Rousing minds to life: Teaching, learning, and schooling in social context*. Cambridge, MA: Cambridge University Press.

Vygotsky, L. S. (1978). *Mind and society: The development of higher mental processes.* Cambridge, MA: Harvard University Press.

Waterman, S. S. (2010). *Assessing middle and high school mathematics and science: Differentiating formative assessment*. Larchmont, NY: Eye On Education.

Wolters, C. A., & Rosenthal, H. (2000). The relation between students' motivational beliefs and their use of motivational regulation strategies'. *International Journal of Educational Research*, 33, 801-820.

Appendix

Figures and charts for two-way ANOVA (Geometry Score)

Between-Subjects Factors

		Value Label	N
Group	0	Grade Only	30
	1	Comments Only	30
	2	Grades and Comments	30
Performance	0	Low	39
	1	High	51

Descriptive Statistics

Dependent Variable: Total Geometry Quiz

Group	Performance	Mean	Std. Deviation	N
Grade Only	Low	3.15	1.772	13
	High	5.47	1.700	17
	Total	4.47	2.063	30
Comments Only	Low	5.85	1.573	13
	High	5.82	1.334	17

	Total	5.83	1.416	30
Grades and Comments	Low	2.77	1.833	13
	High	5.59	1.906	17
	Total	4.37	2.327	30
Total	Low	3.92	2.181	39
	High	5.63	1.637	51
	Total	4.89	2.063	90

Tests of Normality

	Performance	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Total Geometry Quiz	Low	.104	39	.200[*]	.964	39	.233
	High	.133	51	.026	.939	51	.012

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Total Geometry Quiz

F	df1	df2	Sig.
.316	5	84	.902

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Group + Performance + Group * Performance

Tests of Between-Subjects Effects

Dependent Variable: Total Geometry Quiz

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	138.373[a]	5	27.675	9.665	.000	.365	48.327	1.000
Intercept	2015.798	1	2015.798	704.016	.000	.893	704.016	1.000
Group	49.886	2	24.943	8.711	.000	.172	17.423	.965
Performance	64.198	1	64.198	22.421	.000	.211	22.421	.997
Group * Performance	33.886	2	16.943	5.917	.004	.123	11.835	.866
Error	240.516	84	2.863					
Total	2530.000	90						
Corrected Total	378.889	89						

a. R Squared = .365 (Adjusted R Squared = .327)

b. Computed using alpha = .05

Parameter Estimates

Dependent Variable: Total Geometry Quiz

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	Partial Eta Squared	Noncent. Parameter	Observed Power[b]

					Lower Bound	Upper Bound			
Intercept	5.588	.410	13.617	.000	4.772	6.404	.688	13.617	1.000
[Group=0]	-.118	.580	-.203	.840	-1.272	1.037	.000	.203	.055
[Group=1]	.235	.580	.405	.686	-.919	1.389	.002	.405	.069
[Group=2]	0[a]	.	.	.	.	.	.	.	.
[Performance=0]	-2.819	.623	-4.522	.000	-4.059	-1.579	.196	4.522	.994
[Performance=1]	0[a]	.	.	.	.	.	.	.	.
[Group=0] * [Performance=0]	.502	.882	.570	.570	-1.251	2.256	.004	.570	.087
[Group=0] * [Performance=1]	0[a]	.	.	.	.	.	.	.	.
[Group=1] * [Performance=0]	2.842	.882	3.223	.002	1.088	4.595	.110	3.223	.890
[Group=1] * [Performance=1]	0[a]	.	.	.	.	.	.	.	.
[Group=2] * [Performance=0]	0[a]	.	.	.	.	.	.	.	.
[Group=2] * [Performance=1]	0[a]	.	.	.	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

b. Computed using alpha = .05

Pairwise Comparisons

Dependent Variable: Total Geometry Quiz

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
Grade Only	Comments Only	-1.523*	.441	.003	-2.600	-.446
	Grades and Comments	.133	.441	1.000	-.943	1.210
Comments Only	Grade Only	1.523*	.441	.003	.446	2.600
	Grades and Comments	1.656*	.441	.001	.579	2.733
Grades and Comments	Grade Only	-.133	.441	1.000	-1.210	.943
	Comments Only	-1.656*	.441	.001	-2.733	-.579

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Multiple Comparisons

Dependent Variable: Total Geometry Quiz

LSD

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval

					Lower Bound	Upper Bound
Grade Only	Comments Only	-1.37*	.437	.002	-2.24	-.50
	Grades and Comments	.10	.437	.820	-.77	.97
Comments Only	Grade Only	1.37*	.437	.002	.50	2.24
	Grades and Comments	1.47*	.437	.001	.60	2.34
Grades and Comments	Grade Only	-.10	.437	.820	-.97	.77
	Comments Only	-1.47*	.437	.001	-2.34	-.60

Based on observed means.

The error term is Mean Square(Error) = 2.863.

*. The mean difference is significant at the .05 level.

Pairwise Comparisons

Dependent Variable: Total Geometry Quiz

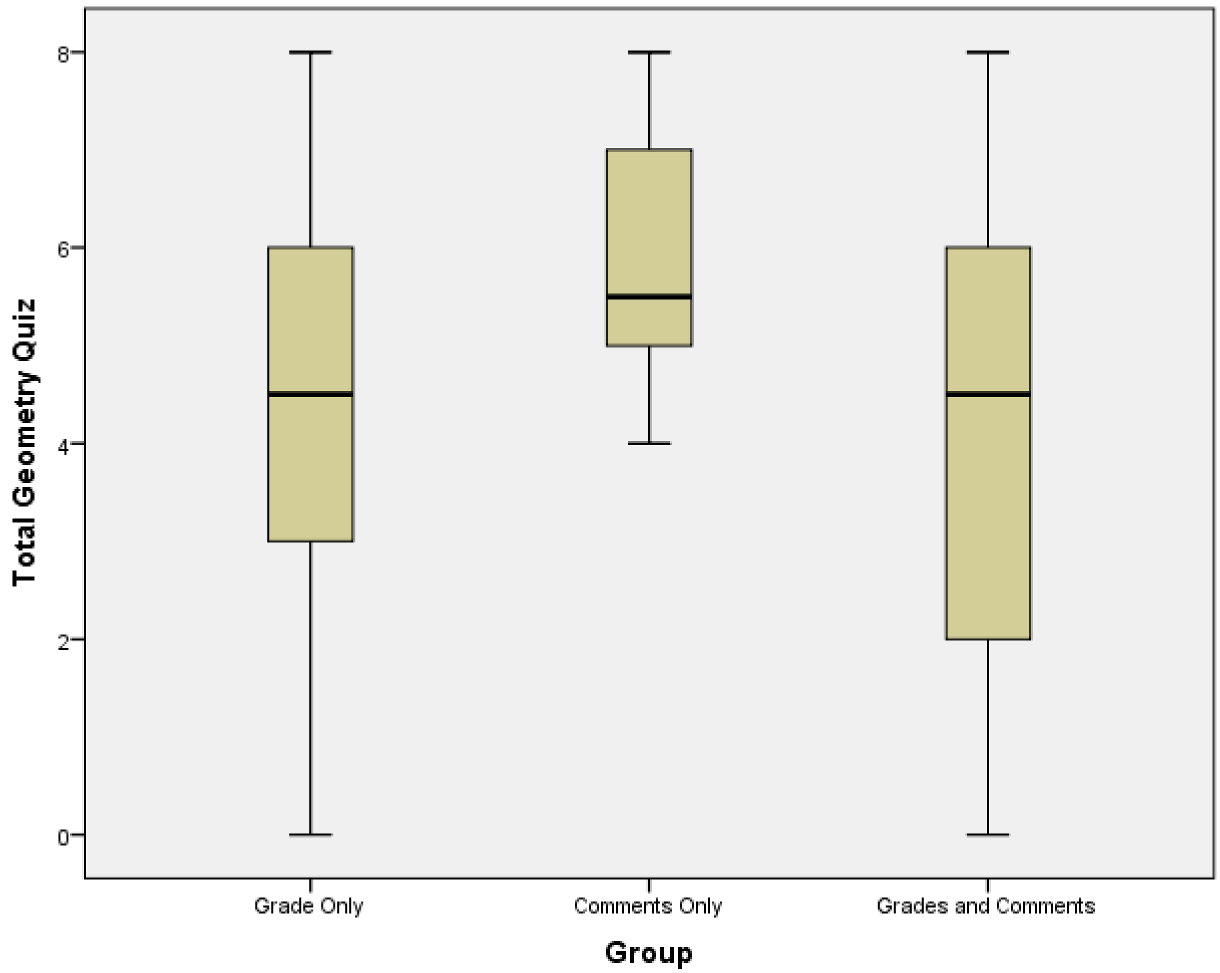
Group	(I) Performance	(J) Performance	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
						Lower Bound	Upper Bound
Grade Only	Low	High	-2.317*	.623	.000364	-3.557	-1.077
	High	Low	2.317*	.623	.000364	1.077	3.557

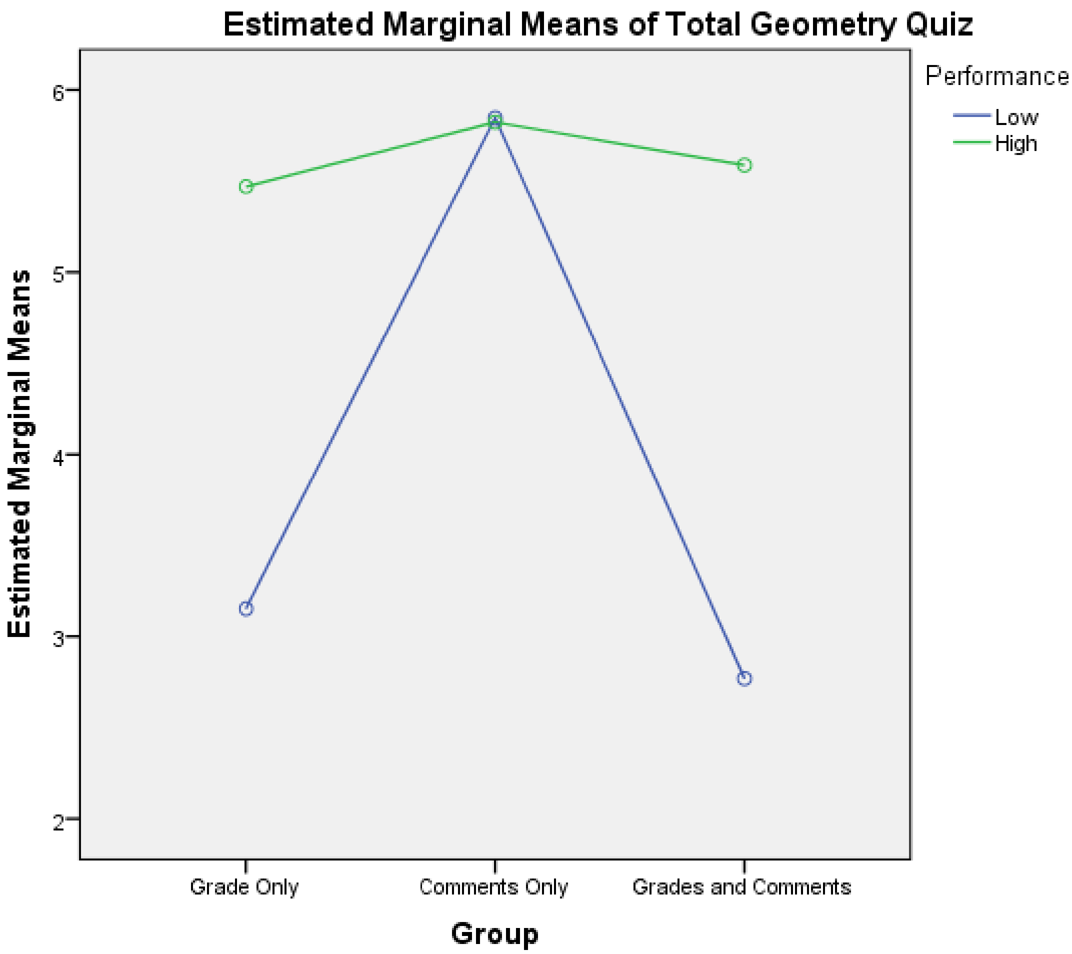
Comments Only	Low	High	.023	.623	.971138	-1.217	1.262
	High	Low	-.023	.623	.971138	-1.262	1.217
Grades and Comments	Low	High	-2.819*	.623	.000020	-4.059	-1.579
	High	Low	2.819*	.623	.000020	1.579	4.059

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).


8
6
4
2
0
Total Geometry Quiz
Grade Only
Comments Only
Grades and Comments
Group


Estimated Marginal Means of Total Geometry Quiz
Performance
Low
High
Estimated Marginal Means
6
5
4
3
2
Grade Only
Comments Only
Grades and Comments
Group

TWO WAY ANOVA /T-Test (Question #8 Geometry)

Tests of Between-Subjects Effects

Dependent Variable: Geometry Question 8

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	3.795[a]	3	1.265	7.404	.000	.205	22.211	.982
Intercept	41.817	1	41.817	244.746	.000	.740	244.746	1.000
Group	1.089	2	.544	3.187	.046	.069	6.373	.596
Performance	2.706	1	2.706	15.838	.000	.156	15.838	.976
Error	14.694	86	.171					
Total	64.000	90						
Corrected Total	18.489	89						

a. R Squared = .205 (Adjusted R Squared = .178)

b. Computed using alpha = .05

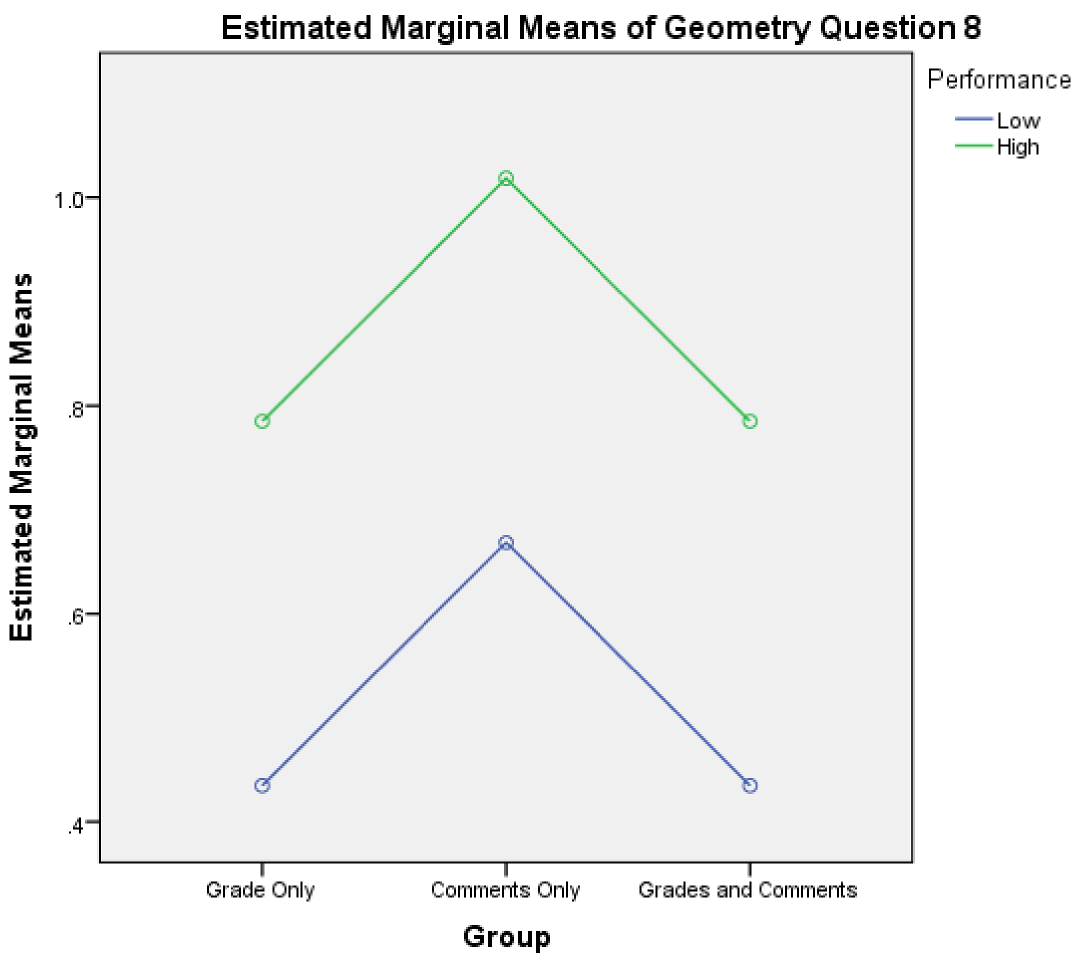
Levene's Test of Equality of Error Variances[a]

Dependent Variable: Geometry Question 8

F	df1	df2	Sig.
10.385	5	84	.000

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Group + Performance



Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
Geometry Question 8 * Group	90	100.0%	0	0.0%	90	100.0%

Report

Geometry Question 8

Group	Mean	N	Std. Deviation
Grade Only	.63	30	.490
Comments Only	.87	30	.346
Grades and Comments	.63	30	.490
Total	.71	90	.456

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
Geometry Question 8	Grade Only	30	.63	.490	.089
	Comments Only	30	.87	.346	.063

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Geometry Question 8	Equal variances assumed	20.073	.000	-2.131	58	.037	-.233	.110	-.453	-.014
	Equal variances not assumed			-2.131	52.133	.038	-.233	.110	-.453	-.014

TWO WAY ANOVA FOR TOTAL QUESTIONARRIE

Between-Subjects Factors

		Value Label	N
Group	0	Grade Only	30
	1	Comments Only	30
	2	Grades and Comments	30
Performance	0	Low	39
	1	High	51

Descriptive Statistics

Dependent Variable: Total Survey

Group	Performance	Mean	Std. Deviation	N
Grade Only	Low	27.92	7.147	13
	High	28.76	7.076	17
	Total	28.40	6.996	30
Comments Only	Low	31.62	6.049	13
	High	33.41	4.651	17
	Total	32.63	5.282	30
Grades and Comments	Low	27.92	7.147	13
	High	28.88	6.431	17
	Total	28.47	6.647	30
Total	Low	29.15	6.850	39
	High	30.35	6.399	51
	Total	29.83	6.588	90

Tests of Between-Subjects Effects

Dependent Variable: Total Survey

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	384.643[a]	3	128.214	3.170	.028	.100	9.511	.716
Intercept	78257.376	1	78257.376	1935.138	.000	.957	1935.138	1.000

Group	352.867	2	176.433	4.363	.016	.092	8.726	.741
Performance	31.776	1	31.776	.786	.378	.009	.786	.142
Error	3477.857	86	40.440					
Total	83965.000	90						
Corrected Total	3862.500	89						

a. R Squared = .100 (Adjusted R Squared = .068)

b. Computed using alpha = .05

Parameter Estimates

Dependent Variable: Total Survey

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval		Partial Eta Squared	Noncent. Parameter	Observed Power[b]
					Lower Bound	Upper Bound			
Intercept	28.986	1.301	22.286	.000	26.401	31.572	.852	22.286	1.000
[Group=0]	-.067	1.642	-.041	.968	-3.331	3.197	.000	.041	.050
[Group=1]	4.167	1.642	2.538	.013	.903	7.431	.070	2.538	.709
[Group=2]	0[a]	.	.	.	.	.	.	.	.
[Performance=0]	-1.199	1.353	-.886	.378	-3.888	1.490	.009	.886	.142
[Performance=1]	0[a]	.	.	.	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

b. Computed using alpha = .05

Lack of Fit Tests

Dependent Variable: Total Survey

Source	Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Lack of Fit	3.993	2	1.997	.048	.953	.001	.097	.057
Pure Error	3473.864	84	41.356					

a. Computed using alpha = .05

Pairwise Comparisons

Dependent Variable: Total Survey

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
Grade Only	Comments Only	-4.233*	1.642	.035	-8.243	-.224
	Grades and Comments	-.067	1.642	1.000	-4.076	3.943
Comments Only	Grade Only	4.233*	1.642	.035	.224	8.243
	Grades and Comments	4.167*	1.642	.039	.157	8.176
Grades and Comments	Grade Only	.067	1.642	1.000	-3.943	4.076

	Comments Only	-4.167*	1.642	.039	-8.176	-.157

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Multiple Comparisons

Dependent Variable: Total Survey

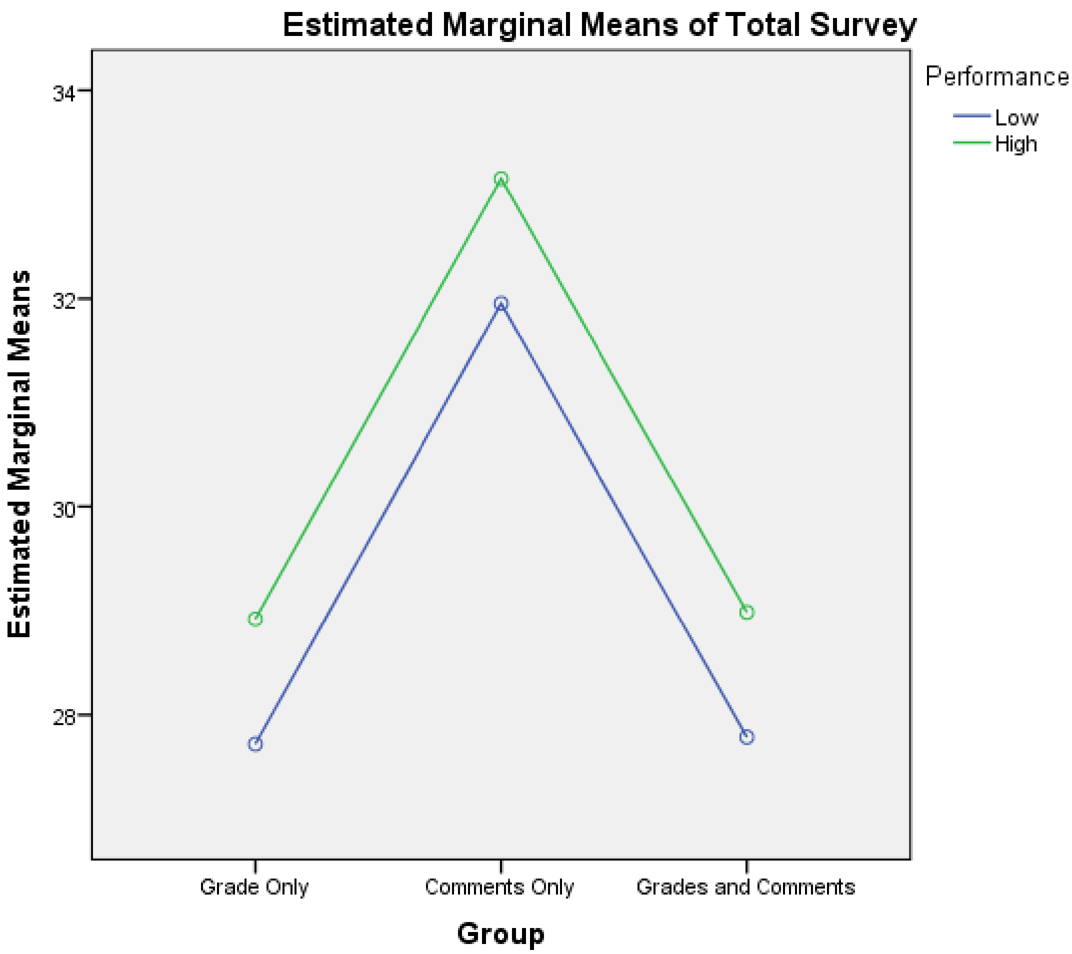
LSD

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Grade Only	Comments Only	-4.23*	1.642	.012	-7.50	-.97
	Grades and Comments	-.07	1.642	.968	-3.33	3.20
Comments Only	Grade Only	4.23*	1.642	.012	.97	7.50
	Grades and Comments	4.17*	1.642	.013	.90	7.43
Grades and Comments	Grade Only	.07	1.642	.968	-3.20	3.33
	Comments Only	-4.17*	1.642	.013	-7.43	-.90

Based on observed means.

The error term is Mean Square(Error) = 40.440.

*. The mean difference is significant at the .05 level.



ANOVA Clicks on Extra Stuff button

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Clicks Under Extra Button

F	df1	df2	Sig.
7.172	5	84	.000

Tests the null hypothesis that the error variance
of the dependent variable is equal across groups.

a. Design: Intercept + Group + Performance

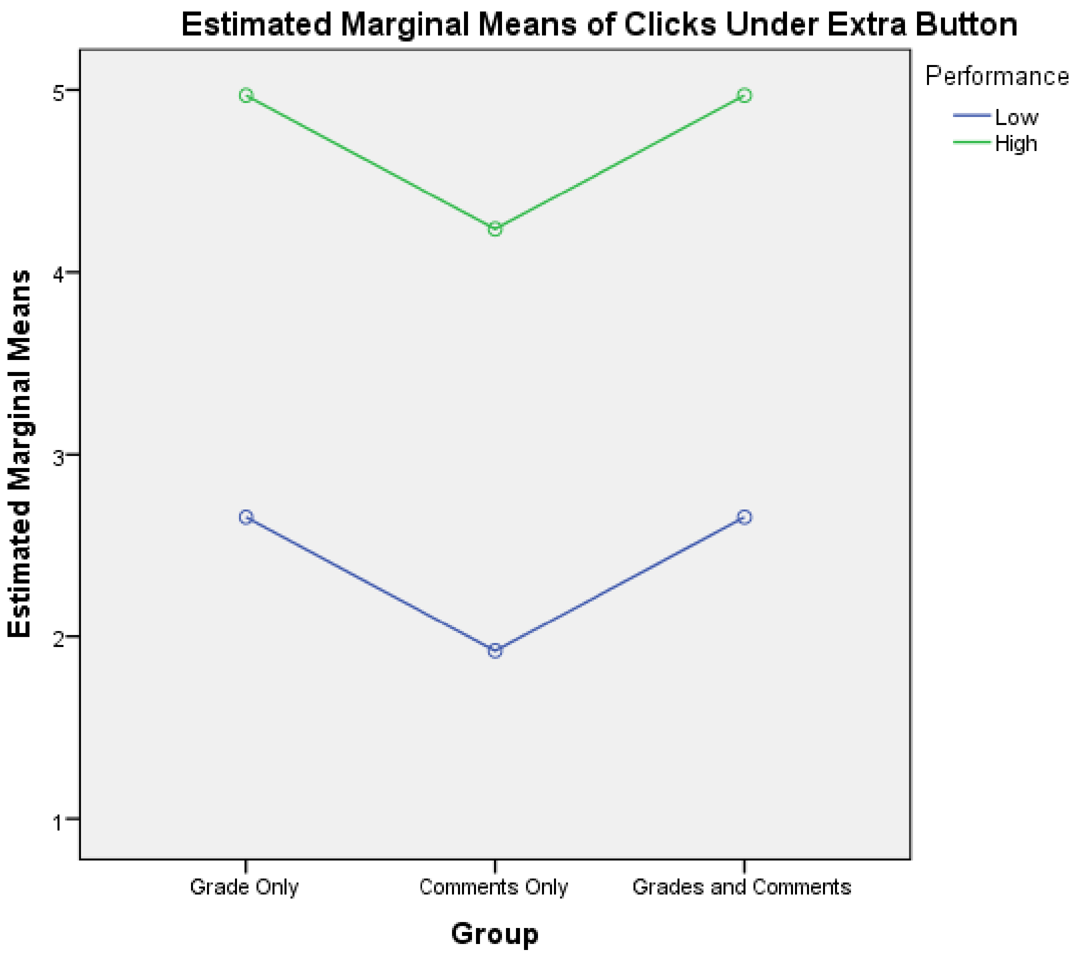
Tests of Between-Subjects Effects

Dependent Variable: Clicks Under Extra Button

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	129.218[a]	3	43.073	2.759	.047	.088	8.276	.648
Intercept	1125.307	1	1125.307	72.069	.000	.456	72.069	1.000
Group	10.756	2	5.378	.344	.710	.008	.689	.104
Performance	118.463	1	118.463	7.587	.007	.081	7.587	.777
Error	1342.837	86	15.614					
Total	2719.000	90						
Corrected Total	1472.056	89						

a. R Squared = .088 (Adjusted R Squared = .056)

b. Computed using alpha = .05


Estimated Marginal Means of Clicks Under Extra Button
Performance
Low
High
Estimated Marginal Means
5
4
3
2
1
Grade Only
Comments Only
Grades and Comments
Group