------------------------------------------------------------------------------------------------------------------------------------------

**Referee: 2**

**Recommendation: Minor Revision**

------------------------------------------------------------------------------------------------------------------------------------------

**Comments:** Thanks for taking the time and putting forth the effort to revise your manuscript, and to address the (I admit, many) comments and suggestions with which I regaled you about the previous draft. I enjoyed reading this revision. Below I include a number of comments on this revision, comments on the changes made to the previous version, and some other minor points.

**Comments:**

- p.1, line 39: I'm not terribly happy with the phrase "Our results suggest." It seems more accurate to say "One possible explanation for our results is." For example, an alternate explanation would be that the study skills of high ability students are such that preparing for the quizzes that replaced the homework was sufficient for them to master the material in the same way that (or slightly better than) homework would have. This wouldn't necessarily be "minimal out-of-class practice"; it would just reflect a difference set of learning skills.

Corrected as recommended.

- p.2, line 10: I would insert "for students" between "enough opportunities" and "to achieve subject mastery," but this is subjective; take it or leave it as you like.   
Corrected as recommended.

- p.3, line 23: I don't think you need to include the numeric reference [25] here, given that it appears above, but, again, this is a subjective point.

Corrected as recommended.

- p.5, line 35--36: My comment about the phrase "Our results suggest" applies here as well.   
Corrected as recommended.

- p.5, regarding students' willingness to work on homework and the number of points it is worth: I think the revised draft reads well. I am interested in your explanation and am glad to have had the chance to read it, but I think your assessment that it does not contribute anything essential to your conclusion is fair.

No revisions required.

- p.6, line 19: Is "within USMA, including the Department" better than the current language?   
Corrected as recommended.

- p.8, lines 19--25: I appreciate your clarification of how your groups were chosen. I'm curious what the actual mean math SAT or ACT for the different groups was; that is, how different are these groups, really? Are the differences statistically significant?

You raise an interesting question that we did not consider previously. Upon closer inspection, we have confirmed a statistically significant difference between groups.

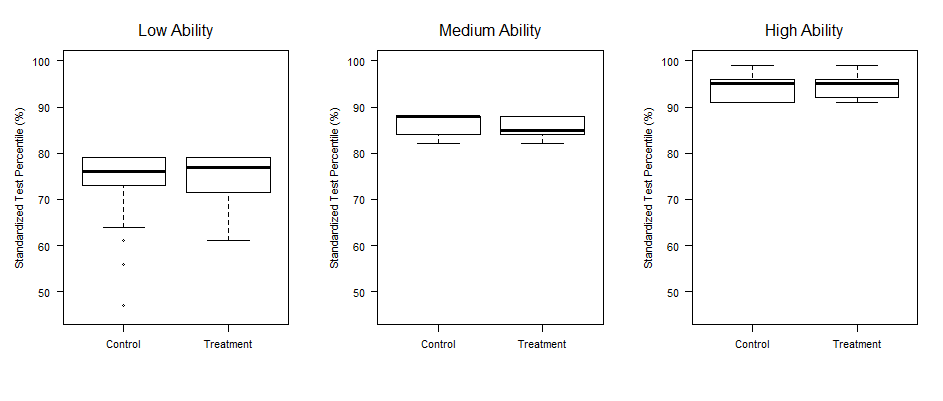
Two sample t-tests for the control group show differences in mean standardized test percentile scores for each group were statistically significant:

* The difference between the high () and medium () ability groups was statistically significant, .
* The difference between the medium () and low () ability groups was statistically significant, .

Two sample t-tests for the treatment group show differences in mean standardized test percentile scores for each group were statistically significant:

* The difference between the high () and medium () ability groups was statistically significant, .
* The difference between the medium () and low () ability groups was statistically significant, .

Last, a visual inspection of box plots illustrates the differences between ability groups and the similarity of ability group performances across the control and treatment groups.



- p.10, lines 20--23: This is probably not within the scope of this paper, but I wonder if the lack of observed impact from homework is for the lower ability students a reflection of their (relative) lack of success on the homework in the first place. That is, if one had a large enough sample that one could restricted consideration only to those students in each group who scored, say, 90% or better on the homework, would the results change?

We agree the lack of observed impact for lower ability students is likely a reflection of their lack of success on the homework, and our experimental data appears to support this explanation to some degree. Unfortunately, our sample size is insufficient to draw statistically significant conclusions.

Only 1 student in the control group had a homework average above 80%, so we are not able to subset the low-ability groups in the manner you described. However, it is possible to compare 9 students in the treatment group who had a homework average of 80% of better with the low-ability students in the control group. These 9 individuals outperformed the low-ability students in the control by 8.4%, 3.8%, 7.6%, and 5.8% on the 2nd, 3rd, 4th, and final exams respectively, and the effect on the final exam was statistically significant.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Homework < 80% | Homework >= 80% | Effect |  |  |
| Exam 1 | 0.837 (0.07) | 0.802 (0.14) | -0.034 | 0.69 (9.74) | 0.500 |
| Exam 2 | 0.757 (0.09) | 0.842 (0.11) | +0.084 | -2.01 (12.1) | 0.066 |
| Exam 3 | 0.732 (0.15) | 0.770 (0.17) | +0.038 | -0.59 (12.5) | 0.564 |
| Exam 4 | 0.706 (0.13) | 0.783 (0.09) | +0.076 | -1.84 (18.5) | 0.080 |
| Final | 0.759 (0.07) | 0.817 (0.05) | +0.058 | -2.33 (18.3) | 0.030\* |

The average performance of the 9 individuals across all exams was only 2.7% higher than the rest of the group and was not statistically significant, .

We agree the lack of observed impact for lower ability students is likely a reflection of their lack of success on the homework, and our experimental data supports this explanation. Only 1 student in the control group had a homework average above 80%, so we are not able to subset the low-ability groups in the manner you described. However, it is possible to compare 9 students in the treatment group who had a homework average of 80% of better with the low-ability students in the control group. A non-parametric test reveals that there is a statistically significant improvement (p-value of .038) in exam scores for students in the low group that had success in their homework. Extending this to the medium and high ability groups in the control, the significance becomes even greater (p-value of .0000126) that success on homework is indicative of success on exams.

Although these findings are interesting, the purpose of our paper is to assess the effect of incentivized homework on student achievement. Hence, we feel it is more appropriate examine the performance of the treatment group as a whole in our results section. On the other hand, we do feel it is appropriate to include these results in our discussion and have edited the first paragraph of the discussion section as follows:

“An additional consideration is that anecdotal evidence (in particular, instructors' observations) suggests that low ability students often relied heavily on assistance from others when completing assignments. Unfortunately, over-reliance on others is a consideration that is difficult to account for and one we did not measure. An alternative explanation is that the lack of observed impact from homework for low ability students was possibly a reflection of the students' lack of success on homework. Our experimental data supports this explanation. We found that 9 out of 20 students in the treatment scored 80% or better on homework, and this group significantly outperformed the other 11 students based on a non-parametric two-sample test with a p-value of 0.038. If this same approach is extended to the medium and high ability groups, the significance increases dramatically (p-value of 0.000013). This supports claims from other research noted in our literature review claiming homework completion correlates with improved test scores.

- p. 12: How many students responded to the attitude surveys for each of control/treatment groups?   
Every student responded to the survey. For clarity, the first sentence of section 3.6 now states “all of the students…” as opposed to “students”.

- p.13, regarding the statistically insignificant findings: I think this is fine.  Thanks for the explanation of the nature of the anecdotal evidence for weaker students' reliance on support.  I think it might help to clarify this, as suggested below.

p.13, line 42/43: perhaps change "...anecdotal evidence suggests that low ability students.." to "anecdotal evidence (in particular, instructors' observations) suggests that low ability students..."

Corrected as recommended.

- p.14, last paragraph: I'm returning to my comment about what conclusion is most appropriate for this study. I wonder if a conclusion is "we should have a variety of assignment types," in light of my proposed alternate hypothesis for the cause of your results.

We appreciate your feedback on our conclusion and have edited our final paragraph to read as follows:

“In summary, our work implies the correlation between incentivized homework and exam performance depends on a student's mathematical ability. Hence, educators should consider a variety of assignment types when determining the most effective way to improve student achievement. We found that medium ability students appear to benefit statistically from incentivizing daily homework. The same was not true for low and high ability students. For low ability students, incentivizing homework may not translate to increased performance on exams. Likewise, the study skills of high ability students may be such that preparing for the quizzes (or other alternative assessments) is sufficient or more effective than homework for mastering course material.”

------------------------------------------------------------------------------------------------------------------------------------------

**Brief Synopsis:** This is a manuscript in the scholarship of teaching and learning, describing a study that sought to determine the effects of providing students graded incentives to complete on-line homework. The authors report that such incentives significantly increase the fraction of students completing the homework, and students they classify as medium ability appear to benefit from this in their exam scores. They see little or no statistically significant effects for other students, nor in what is defined as long-term knowledge acquisition by students.

**Audience:** The audience for this manuscript is instructors of college-level mathematics and those with an interest in research on the implementation and effect of pedagogical strategies such as on-line homework. It is well structured for those audiences.

**Exposition:** The overall writing in the paper is fine. It is a good length, and generally reads well. I include some specific suggestions for improvement in the comments to the authors, below.

**Mathematical Content:** The mathematical content appears good.  I appreciate the authors' explanation of their statistical grouping as it is provided in the response letter.  I am curious whether the authors' results hold up if the data are grouped first by section, but do not think this needs be included in this paper.

**Professionalism:** The professionalism of the work is fine. The authors demonstrate a good understanding of the literature to which they are adding, and how the current work fits into that. Their work with the editor and reviewers has resulted in a stronger and more interesting manuscript.

**Fit:** This manuscript appears to be a good fit for PRIMUS. It includes "(2) discussion of how [an] idea impacts student learning...", and "(4) exploration of at least one interesting problem in undergraduate mathematics pedagogy, as well of substantive insight about that problem..."

**Overall Merit:** I think the manuscript as revised should be published in PRIMUS, subject to minor revisions suggested above in my comments to the authors.

------------------------------------------------------------------------------------------------------------------------------------------

**Associate Editor Comments to the Author:**

------------------------------------------------------------------------------------------------------------------------------------------

Thank you for this good revision to the original manuscript. I am recommending that we accept the paper subject to minor revisions that will not require additional refereeing.

First, please address all of the comments of the referee. I especially agree with this assessment: "I'm not terribly happy with the phrase 'Our results suggest.'" and concur that phrasing like "One possible interpretation of our results is ..." Please follow the corresponding recommendations to address this phrasing in several locations. The other referee suggestions are important, too.

Second, in reading the paper a second time, I was struck differently by the phrases "high ability", "medium ability", and "low ability" (particularly the last one: "low ability"). I know that you don't mean the students have "low ability to ultimately be successful", but I can also see how some people will read it that way. I recognize that you do explain what you mean by the term and how it is connected to performance on standardized tests. I think you'd be well-served to address this as early as possible in the paper, saying something like "We use the term 'ability' as a shorthand for 'performance on a standardized test'. We of course recognize that a single score on a standardized test is not a summary of a student's overall ability to be successful." I don't think you need to belabor this point, but again think that since the terms are being used in the abstract and early in the paper, it would serve you well to clarify.

Again, we fundamentally like the paper and look forward to seeing it published in the journal subject to these minor revisions. Thank you for using PRIMUS as an outlet for your work.

We are excited to hear your decision to conditional accept our manuscript for publication and greatly appreciate all of the referee feedback. We feel our paper has greatly improved from the initial submission and have done our best to incorporate all of the feedback we received in the second review.

We have provided a point-by-point response for referee 2’s suggestions and also incorporated your recommendation to clarify our use of the phrase “ability”.