EFFECTS OF STUDENT SELF-ASSESSMENT IN AN ENGINEERING CLASSROOM

by

Dylan James Nelson

A professional paper submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Science Education

MONTANA STATE UNIVERSITY
Bozeman, Montana

July 2021

ACKNOWLEDGEMENTS

There are quite a few people that I need to formally thank for their help and support throughout this project. First, I wish to thank Adam Dean, my reader, for giving his time and advice to help me throughout this project. It has been most helpful. I wish to thank Kayelynn Warner for her advice and editorial skills on earlier drafts of this project. I also wish to thank Walt Woolbaugh for his tireless help and guidance. Lastly, I wish to thank my dear wife, Brianna Nelson, for the support she has offered me throughout this whole process and for her wonderful editorial skills.

TABLE OF CONTENTS

TABLE OF CONTENTS CONT.

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Self-assessment has had much research done in the past, but has not been tested for its effectiveness in the classroom by the author. This study details the effectiveness of self-assessment on student achievement in an engineering classroom. Two middle school Pre-Engineering Technology (P.E.T) classes from a Title I school participated in the study with each class being split up into a treatment group and comparison group. All students were given pre-content, unit content, and post-content assessments with the treatment group self-assessing during each unit content assessment. All students participated in Likert Scale Surveys and students were picked at random to participate in interviews. After analysis of data, student self-assessment was found to positively affect student achievement and student confidence, in many cases, when defining engineering problems. The analysis of student self-assessment also proved to be beneficial to the author's teaching practices.

INTRODUCTION AND BACKGROUND

Context of the Study

The topic of this research project was on self-assessment and its effects in the classroom. According to McDonald (2013), self-assessment is defined as, "the involvement of students in identifying standards and/or criteria to apply to their work and making judgments about the extent to which they have met these criteria and standards" (p. 62). As this topic has been highly tested in the past, it has been dabbled with by the author in past classes: Coding and Pre-Engineering Technology (P.E.T). After having experimented with it for a year, no clear effect on student achievement was found. Some students seemed to self-assess quite well and were able to meet desired standards (and know why they were able to meet the standards), while others were not able to self-assess accurately and did not meet desired standards. Thus, the question of, "How effective is this in the classroom?" was often asked. When the author started investigating topics as a source for this study, the topic that emerged to the forefront of all other options became the subject of student self-assessment and its effectiveness on student achievement.

This topic has great significance, particularly to the author. If it is a fruitful practice, then it can be used to better help students reach their goals and specific standards while helping the author become a better teacher. If the practice does not turn out fruitful, then efforts can be focused on other practices that will better help students reach goals and standards.

This subject of self-assessment and the results that come from this study may also be of significance to other people/groups apart from the author. Though some students do not currently understand the potential magnitude that self-assessing may have on their achievement in class, this subject and study may be of great importance to them as it may give them insight in how to be successful in current and future classes. Similarly, the administration at the current middle school that this study took place in understands the potential impact that self-assessment can have on student learning, and therefore the results of this study may have some influence on how they work with teachers. Additionally, educators known by the author are interested in the results of the present study.

Of the author, particular interest has been on how effective student self-assessment is in his P.E.T classes, which are classes that would fall under the engineering or STEM (Science, Technology, Engineering, Mathematics) category. Students only get this class for a trimester out of the year, so time with them is quite short compared to other teachers. As well, only four engineering standards are used to assess student progress. These standards are run through each class in a cyclical manner (meaning, there are about three units that are covered each trimester, and each standard is covered in each unit, therefore making the curriculum a cyclical or spiral curriculum).

One of the standards covered in this class is MS-ETS1-1. As dictated by NGSS Lead States (2013), this standard sets forth that students ought to be able to, “Define the criteria and constraints of a design problem with sufficient precision to ensure a successful solution” (p. 244). As an engineer cannot solve a problem without defining the problem first, neither can engineering students solve unit problems without proficiently defining the problem first. This standard permeates through all of the units taught and sets students up to be able to create successful solutions to problems (if they have defined the problem accurately enough).

The primary purpose of this study was to find what effects self-assessment has on student achievement when defining an engineering problem. As self-assessment may have other effects in the classroom, and the fact that the nature of this research will be action research that will benefit the author and his students, there are also secondary purposes to this study.

Focus Question

My focus question was, What effect does student self-assessment have on student achievement when defining engineering problems?

My sub-questions include the following:

1. What effect does student self-assessment have on student confidence when defining an engineering problem?
2. How does the analysis of student self-assessment improve my teaching?

CONCEPTUAL FRAMEWORK

Much previous research has been performed about self-assessment. In the following sections, analysis and synthesis of the research on, and basis of, self-assessment will be reviewed. Information has been gathered from various peer-reviewed journal articles, an article from a conference, and two books. The themes and findings related to this study have been outlined in the following sections.

Underlying Self-Assessment: Constructivism

It seems that the goal of most (if not all) practices in education are to enable a student to learn. According to Ormrod (2012), learning is a "change in mental representations of associations due to experience" (p. 18). According to this definition, learning happens through experience. As this is very much the case in many, if not all, contexts, it can be said that a person learns through constructing their perspectives based on their experiences and observations (Ormrod, 2012). This is the basis of constructivism.

Self-assessment is a sub-set of constructivism. On the subject of self-assessment, Martin et al. (2017) implies that, when learning, a person builds their knowledge through the activities that are performed, knowledge is built and assessed through self-assessment. Therefore, as Adeyemi (2012) puts it, when a person assesses their own learning, they are building knowledge in that area through identifying needs, setting goals, and monitoring progress in relation to the achievement of these goals.

Through the constructivist lens, observation and theory would suggest that a person cannot construct their own learning unless they were motivated toward and engaged in learning observations and experiences. Just as a dog will not engage in fetching a ball until a stimulus is provided or realized, a person will not construct learning in their lives unless a stimulus (whether external or internal) provides this jump. Self-assessment is reported to offer a stimulus to learners that can increase a learner's motivation through engaging a learner as an active participant in learning (Adeyemi, 2012). Bourke (2016) states that as students actively participate in their learning, "the outcomes tend to be improved…motivation and…engagement," which can eventually, "[increase] student achievement" (p. 100). Along with engagement, self-assessment has been reported to also empower students in their ability to build their learning (McDonald, 2013), according to McDonald and Boud (2003) it can influence students to take responsibility of their learning, and, based on what Semana and Santos (2018) have found, it can help students judge their steps toward learning to help monitor how they are building their knowledge.

Truly, in regard to the construction of knowledge, self-assessment is at the core of this practice.

Self-Assessment in the Classroom

Though self-assessment has been reported to have a positive impact on student learning (McDonald & Boud, 2003), it has been noted that this impact will only happen if a few key findings are followed.

Studies produced by Elder (2010), Fastre et al. (2012), and Harris and Brown (2013) have reported that when students self-assess, many will often do so inaccurately. Similarly, on the subject of self-assessment, Geringer (2017), has found that students tend to rank themselves higher than they actually achieved. It has also been found that students who don't perform well or don't understand the criteria to be met well enough, will over-rank themselves (Gehringer, 2017), and that, according to Andrade (2019), self-assessment from older students (college level) are often more accurate than younger students (grade-school level, particularly in the younger grades). Much inaccuracy in this type of assessment is often due to not having the self-assessment skill well developed (Gehringer, 2017) because students have had little experience with it (or just the fact that it is completely new to them) (Fastre et al., 2012).

In response to the above inaccuracies, studies have suggested, and found, that specific practices, when followed, will allow for more accurate self-assessment, thus boosting student achievement. One such practice is the practice of formally training students on how to self-assess (McDonald & Boud, 2003; Semana & Santos, 2018). Many students don't have this skill, therefore, formal training on how to use the skill in specific contexts is necessary. Heritage (2009) puts forth that students ought to be given the opportunity to practice self-assessment after being formally trained, which practice will then help self-assessments become more accurate over time (Gehringer, 2017).

As well, if students are helped to understand the criteria by which they will be rating themselves, student accuracy will increase—this is true particularly with lower achieving students (Fastre et al., 2012). Almarode et al. (2018) justify this finding by stating, "when students understand the success criteria, they then can be most involved in assessing their own success, and their progression toward this success" (p. 38). As students more accurately self-assess based on understood criteria, achievement of learning expectations will become more attainable to students, thus increasing student self-efficacy (Adeyemi, 2012). Alerting a student to the criteria through which they will be assessed has been found to be particularly helpful when placed in a rubric (Andrade, 2019).

In regard to teachers, further advice is given to help self-assessment be more accurate and effective in the classroom. One such piece of advice is that self-assessment ought to be used primarily as more of a formative assessment (Andrade, 2019; Heritage, 2009), rather than as a summative assessment (Gehringer, 2017). If students think that the self-assessment will hurt their grade in any way, many will likely over-rate themselves (Gehringer, 2017). Thus, if the assessment is purely formative in regard to the progress of their learning, they will be more likely to be honest about their own achievement. Self-assessment will also be more effective if the teacher purposely plans the self-assessment activities as part of the curriculum and uses the assessments to guide their instruction (McDonald, 2013).

Methodologies of Past Studies

As the subject of self-assessment has been studied and researched time and again, there are a few valuable items to be drawn from the methodologies of these studies. A few studies were able to draw from large populations of students for their studies. One drew from a base of 515 students (McDonald, 2013) and another from 10 high schools with a population totaling over 1000 students (Mcdonald & Boud, p. 2003). In other studies, whether because of the way the study was set up or because of the lack of people available for the study, whole populations within classrooms were used (Fastre et al., 2012; Martin et al., 2017). For most circumstances, and particularly for the circumstance of this study, using the whole classroom population was most doable and was the most likely set-up to be enable the most reliable data possible.

Of the populations in the studies reviewed, most utilized populations that represented a breadth of students ranging from those who had high achievement in school to those who had low achievement (Elder, 2010; McDonald, 2013; McDonald & Boud, 2003). The study conducted by Harris and Brown (2013) drew from populations across the grade school levels (primary to secondary grades) and used all students in those classes to represent the breadth of achievement that is so common in classrooms everywhere.

One study was purely, or mostly, qualitative (Martin et al., 2017), one seemed to be mostly quantitative (Adeyemi, 2012), but most seemed to have a mixed method approach to them. The mixed method approach, at least when studying the effects of self-assessment on classroom achievement and student attitudes, seems best. As Semana and Santos (2018) put it, it is best to use a mixed method approach in this instance so researchers can, "improve explanations and enhance the interpretation process" (p. 745).

Regarding this mixed method approach, many studies utilized questionnaires or surveys (Elder, 2010; McDonald, 2013; Semana & Santos, 2018) and a few used interviews to gather data (Elder, 2010; Harris & Brown, 2013; Semana & Santos, 2018).

When using questionaries or surveys, the practice of using Likert scales or Likert items became a reasonable way to gather data and analyze that data through qualitative and quantitative means (Elder, 2010). Other valuable data was further gathered with artifacts that were created by student's actual written explanations of self-assessments (Harris & Brown, 2013; Martin et al., 2017).

Main Findings

Constructivist perspectives lie at the core of self-assessment. Because self-assessment is to be used to engage a student in taking charge of the construction of their learning, student motivation can increase which has been seen to increase student achievement.

This increased achievement cannot happen unless certain practices are followed, namely: giving students formal training on how to self-asses, giving students opportunities to practice self-assessment, and using self-assessment as a formative practice. In doing this, self-assessment will become more accurate and a stronger association between self-assessment and learning can be had.

When performing studies on self-assessment, it has been found that many studies gather data from a range of achievement and age groups, along with utilizing a mixed methods approach that utilizes student artifacts, questionnaires or surveys, and interviews.

METHODOLOGY

The purpose of this study was to find how effective the practice of self-assessment was in an engineering classroom. Particularly, the main points of this study were to find how student self-assessment affected student achievement when defining engineering problems, how student self-assessment affected student confidence when defining engineering problems, and how the analysis of student self-assessment improved the author's teaching abilities.

The sections below will delve into the details on the population involved in the study, the treatment used in the study, and the data collection and analysis strategies of the study.

Demographics

As was mentioned earlier, both Pre-Engineering Technology (P.E.T) classes of the author were used as the population for the present study during the first trimester of the 2020-2021 school year. As many students as possible from each class were chosen to be part of the population of this study to help make the study more reliable, just as Elder (2010), McDonald (2013), and McDonald and Boud (2003) mentioned in the previous chapter. As well, because a sixth grade and an eighth grade P.E.T class were taught by the author, students across these two grade levels were participants in the study.

The school that the study took place in, Farnsworth Middle School, is a Title I school. About 34% of the students of the school would have been given free and reduced lunch (due to the Covid-19 pandemic, all students in the school were offered free breakfast and lunch). Many of these students were Hispanic students. A much smaller portion of these students were Polynesian or of Asian descent.

As well, Farnsworth Middle School is a rural school with many of the student's parents being employed in anything from farming and custodial work to large business owners and engineers.

The students in each class used as the population of the study represented a wide range of diversity. In the sixth grade class, two students were considered to be of a minority ethnicity (Hispanic), and the rest were Caucasian. Of the population, four were females (who achieved highly in all their classes) and the remainder were males (16) who had a wide range of achievement (anywhere from failing a few classes to A's in all classes). One student was on an Individualized Education Program (IEP), one was being assessed to be placed on a learning plan, one student was Limited English Proficient (LEP) and was on a plan for that, one was on a health notification plan, and one was in the gifted and talented program.

Of the students in the eighth grade that participated in the study, two were girls (who also achieved highly in their classes), and 12 were boys. Only one student of the population was of a minority population (Hispanic). In the group of those participating in the study, none were on learning plans. The boys participating demonstrated a wide range of achievement that was like the boys in the sixth grade class. Some were failing classes here and there, some had all A's, and some were in between these extremes. The two girls that participated in the study, like the girls in the 6th grade class, excelled in their classes and obtained mostly A's. It must be noted that not all students in this class (the eighth grade class) participated in the study. Many students were absent the first few days of class when data started being gathered for the study because of the Covid-19 pandemic. In the class there was a total of 22 students, but only 14 showed during the first few days of class when data gathering started. Therefore, data from the 8 missing students was not analyzed. Among this group of students, two had IEPs and chronically struggled with achievement in their classes. There were also other low, mid, and high achieving students in this excluded group that may have influenced the data that was gathered had they been part of the study.

The research methodology for this study received an exemption from Montana State University's Institutional Review Board (see Appendix A).

Treatment

Jefferson County School District, located in Rigby, Idaho and its surrounding parts, runs on the trimester system. The present study was conducted on the population of the author's two P.E.T classes through the duration of the first trimester of the 2020-2021 school year (September 2020-November 2020). Of this study, the treatment period ran from the middle of September through the first week of November.

Each class was split into two groups, a treatment group and a comparison group. This was done on the first day of school. All students on the author's class rosters were placed into an automatic seating chart generator to fill up the available seats in the classroom. After all students were placed into a random seat, the classroom was split into half with one side being the students who were part of the treatment group and the other side being students who were part of the comparison group. Though this set up made it so that students were randomized to help the study be as reliable as possible, the groups created were not necessarily "equal" due to the variability of the students.

All students from each class, no matter which group they were a part of, were formally trained on how to fully define an engineering problem (including identifying the criteria and constraints of that problem) and were given practice as most students had not done so before (and those who had forgotten). This practice was done through a mini unit. During this time all students were introduced to how they would be assessed when defining engineering problems through the use of a rubric (Appendix B). This part of the treatment took about four class periods. At that point in time students in the treatment group were not formally taught how to self-assess to avoid giving the students too much information at one time.

This point in time—after the mini-unit was complete (mid September)—is when the treatment (the act of treatment group students self-assessing) of the study started. Each time the classes started a new engineering unit, all students defined the engineering problem that that unit was built around. After defining each engineering problem, students from the treatment group formally self-assessed on how well they felt they defined the problem. Students of the comparison group did not formally self-assess, but instead, participated in an alternate activity that was not related to self-assessment, but yet still had relevance (these students rated the criteria and constraints of the engineering problem from most important to least important). To help make it so that no group felt "more important" than the other group, it was explained that they were doing different activities because the author wanted to find out how effective each activity was in helping them define engineering problems—the activities done in both groups were equally important.

During the first full unit, and after each student had defined the engineering problem of that unit, the students from the treatment group were formally taught how to self-assess themselves based on the rubric they were being evaluated with. Similarly, students from the comparison group were formally taught how to accomplish their separate task so that all knew how to do their task.

Data Collection and Analysis Strategies

In Table 1 below, the details of instrumentation for each of the research questions of this project are presented.

Table 1. Triangulation matrix of data collection methods per each research question.

Question	Data Collection Method 1	Data Collection Method 2	Data Collection Method 3
What effect does student self-assessment have on student achievement when defining engineering problems?	Pre-Content Assessment; Post-Content Assessment; Unit Content Assessments (Quantity: 3)	Student Interviews	Likert Scale Survey
What effect does student self-assessment have on student confidence when defining an engineering problem?	Pre-Content Assessment; Post-Content Assessment; Unit Content Assessments (Quantity: 3)	Student Interviews	Likert Scale Survey
How does the analysis of student self-assessment improve my teaching?	Pre-Content Assessment; Post-Content Assessment; Unit Content Assessments (Quantity: 3)	Teaching Journal	Student Interviews And Likert Scale Survey

The data collection method that permeates through all the associated questions of this research is the implementation of the Pre, Post, and Unit Content Assessments. Each of these assessments consists of an engineering problem prompt and a blank design brief for the intent of students to use to define each engineering problem. The Pre and Post content Assessments are found in Appendix C, the sixth grade Unit Content Assessments are found in Appendix D, and the Unit Content Assessments for eighth grade are found in Appendix E. Accompanied with the Unit Content Assessments was a blank self-assessment form for the treatment group (Appendix F).

Each Unit Content Assessment piece was developed by the researcher in previous years and had been tweaked over time to be more valid in regard to containing necessary criteria and constraints that were identifiable to students. Over time the data from student completions became more comparable, which helped the instruments become more reliable over time through further tweaking of the assessments. Additionally, Walt Woolbaugh, my department chair, also reviewed the assessments to help insure validity and reliability of the assessments.

The Pre/Post Content Assessments were also developed by the author to assess, like the unit assessments, how accurate a student was at defining a design problem and to also be able to assess student growth and to see if there was a difference in growth between the two groups. To assure the validity and reliability of this instrument, it was created to be the same format as the Unit Content Assessments: complete with perceived identifiable criteria and constraints that were individualized for the differences in sixth grade and eighth grade evaluation abilities.

The self-assessment page for the treatment group students to fill out after defining the unit engineering problems was developed by the author as well. It was created in similar fashion to past self-assessment pages that had seemed to be effective to the author. This was also sent to the science reader and one of the support people of this study to determine the perceived validity of the instrument. A positive answer came back in each instance. The page being created in an easy to navigate way also made the instrument reliable as it was able to be student friendly and easy to use.

The Likert Scale Survey (Appendix G) was developed by the author based on many other Likert scale surveys seen throughout other studies. It consists of six statements that require the person filling it out to mark a stance on their level of agreement with each statement. There are also a few places for the person taking the survey to explain why they marked their stance the way they did. This survey was sent to the science reader of this research project, and advice was given back about how to make it more valid for what it was intended to measure and changes were made. The survey is found in Appendix G. This survey is made more reliable as it was given to all students in both sixth grade and eighth grade participating in the study.

The student interview questions were, like the other instruments of this study, developed by the author (Appendix H). The questions of the interview were also sent to the science reader and another helper of the project to assure validity, and a positive response was returned. Reliability for this instrument was also assured by providing these interview questions to both classes that were part of the study.

The last instrument used by the author was a teaching journal to record the author's observations and goings on in the classroom during the study. This is a technique that has been used widely by other researchers and teachers, which helps to establish the validity of this tool. As well, when recording in the journal, the researcher kept thoughts in line with the research questions, which further established the validity of the tool. Many entries were made to ensure the reliability of the tool.

DATA ANALYSIS

Effects of Self-Assessment on Achievement

The first, and primary, research question of this study asks how self-assessment affects student achievement when defining engineering problems. Findings to this question vary based on grade and student abilities. Data that answers this question will first be analyzed looking at sixth and eighth grade data combined, then the individual grade levels will be analyzed.

Sixth and Eighth Grade Combined

Figure 1 (shown below) shows a box and whisker plot for the distribution of the combined scores of the assessments of the treatment and comparison groups in the sixth and eighth grade classes throughout the trimester.

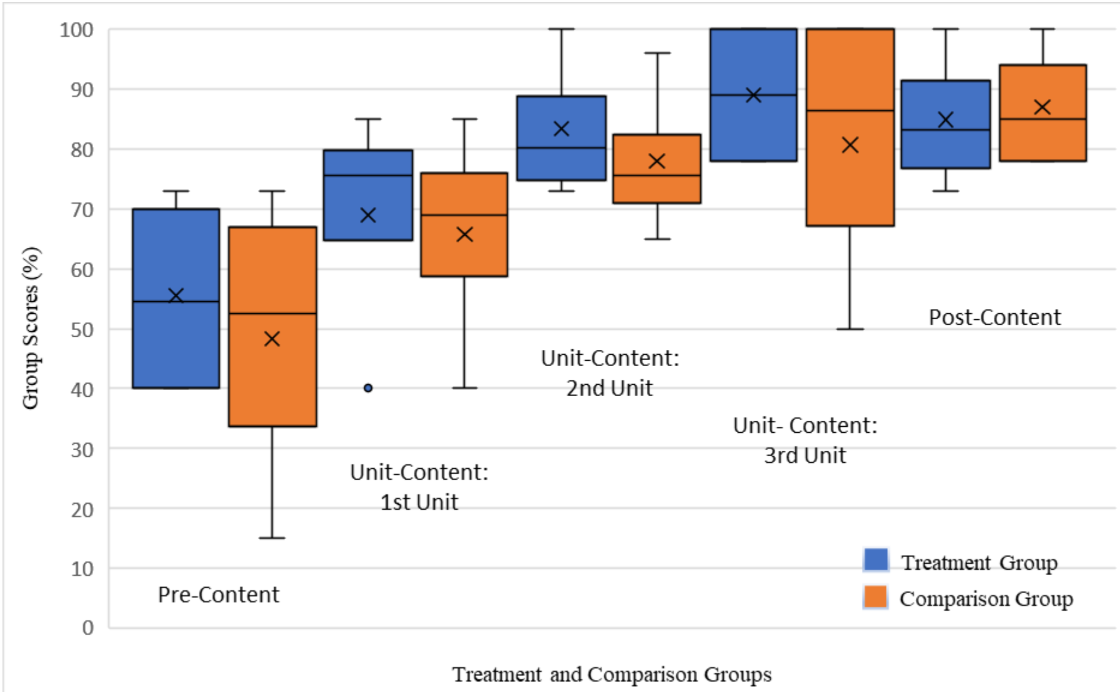


Figure 1. 6th and 8th grade P.E.T treatment group vs. comparison group scores of assessments, (*N*=34).

Based on the figure above, it may be seen that the top 75% of students from the treatment group performed similarly to the top 75% of students from the comparison group on the pre-content assessment. It may also be seen that, starting off, no student was able to achieve higher than a C (a *Basic* score according to the rubric). There is a difference between the lowest 25% percent of students from each group as the lowest 25% of students from the comparison group received lower scores than that of their counterparts. If an independent t-test is conducted to find if there is a statistically significant difference between the starting achievement of the two groups, it is found that if there was no difference between the two groups, the scores achieved on both sides would happen about 69% of the time if there was no difference between the mean achievement of each group. As such, there is no significant statistical difference between the two groups' starting points. This lack of difference likely goes to show that all students were starting off on the same foot with no one really knowing how to define an engineering problem.

If the ranges of scores from the unit assessments from both groups are observed, the range of the treatment group is packed closer together about half of the time (based on the Coefficient of Variation for each groups' score on the different assessments). Despite this lack of difference, the distributions of scores of the treatment group are generally higher achieving than that of the comparison group. The mean scores of both groups were analyzed with independent t-tests to determine if the difference in mean scores was statistically significant enough to provide evidence that self-assessment affects achievement, and the following p-values were obtained (in order of the assessments): 0.32, 0.14, and 0.18. These p-values state that if it is assumed that there was no statistical difference between the achievement of the two groups, the scores achieved would have been 32%, 14%, and 18% likely. None of these p-values fully qualify as providing statistically significant proof that there is a difference between the achievement of the two groups (a number of that magnitude would need to be 5% or lower for studies of this sort), which may be interpreted to mean that student self-assessment didn't affect student achievement in this study. Nonetheless, it is interesting to note that the second and third assessments come close to the 5% threshold, with one being at 14% likely and the other being at 18% likely. As these numbers are close to the 5% threshold, perhaps there may be reason to suggest that the treatment group had higher achievement, at least in part, due to the fact that they self-assessed as this is the only difference in the courses between the two groups.

This claim, that self-assessment may have aided the treatment group students with greater achievement, can be further backed from data generated from student perceptions. When given questions in a Likert-Scale survey that assessed student perceptions of how well they were able to achieve *Proficient* or *Advanced* when defining engineering problems throughout the trimester, there was not much difference between the responses of the two group. For the most part, responses were nearly the exact same. But, when specifically asked about self-assessing and its effects on their abilities to achieve *Proficient* or *Advanced* (B's and A's), student perceptions became clearer. Of the nine students who were interviewed at random, seven stated that self-assessing helped them achieve a *Proficient* or *Advanced* (78%). In explaining why they felt thus, students gave answers similar to the following quote: "Because it helped me think, so my mind-set was on proficient and advanced and it just kind of helped me get to there." On the flip side, when looking at the perceptions of those from the comparison group, 6 out of 9 (67%) of students stated that the alternate activity they did (an activity that didn't involve self-assessment) made no difference to their achievement. Many of these students shared views like the following: "I don't really think it affected what I got." It would seem, based on the interview data obtained, that the likely difference between the achievement of the treatment group in relation to the comparison group, even though the differences are not statistically significant, is due to the fact that the treatment group self-assessed and the comparison group did not.

One last matter or importance to be assessed is what looks to be an apparent digression of the treatment group on the post-assessment. In Figure 1 the treatment group's spread of scores steadily climbs up to greater and greater achievement throughout the trimester until the Post-Assessment, while the comparison group makes a large jump in their achievement. At the post assessment, the ranges of scores of the treatment group fall a decent amount to where the range of scores of the comparison group is slightly better than that of the treatment group (which was not the case for the vast majority of the trimester). It can also be seen that the mean score for the treatment group fell from close to 90% (almost an *Advanced* score) to about an 83% (below a *Proficient* score). Much of this digression stemmed from a confusion between the criteria and constraints of the assessment problem. The cause of the digression is explained below in an excerpt from the author's teaching journal:

> In this [assessment] the vast majority of students got *Basic* [a C] because of a confusion as to what constraints really are in engineering. The issue happened in both grades. In both grades there was a criteria where "Mearl" shouldn't have to lift her feet above five inches. Understandably, many students put this as a constraint, when, in reality, this was a criteria (November 12, 2020).

If the criteria that caused the confusion for the students is omitted from the data, the scores for the trimester would look as is shown in the figure below.

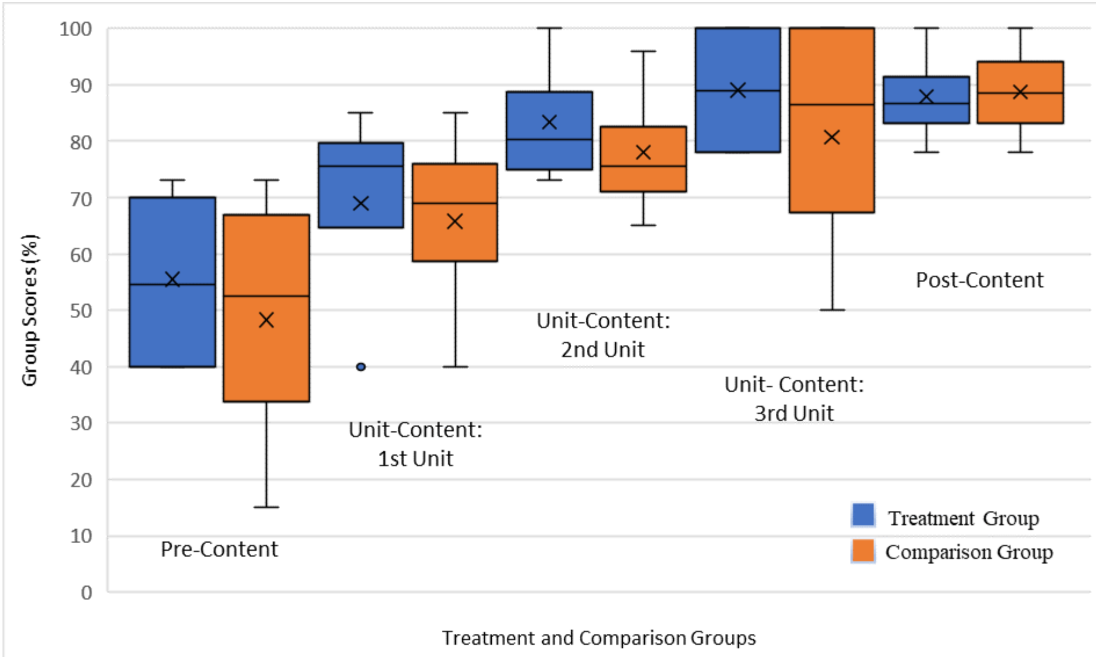


Figure 2. 6th and 8th grade PET scores with adjusted scores to the post-content assessment, (*N*=34)

Treatment group scores go back to about where they were during the third unit assessment, just with scores more closely focused around the proficient range. The scores of the comparison group also improve to be almost exactly the same as the treatment group (and even slightly better). Before the omission, eight to nine of the students from each group were below *Proficient* range, but after the omission only two students from each group were below *Proficient* range, all of which is data to show that all students had a misconception as to what criteria and constraints really are in engineering.

<u>Sixth Grade</u>

Data gathered from participating sixth graders has been compiled in the figure below (Figure 2). The data consists of the range of scores of each group through all the assessments of the trimester.

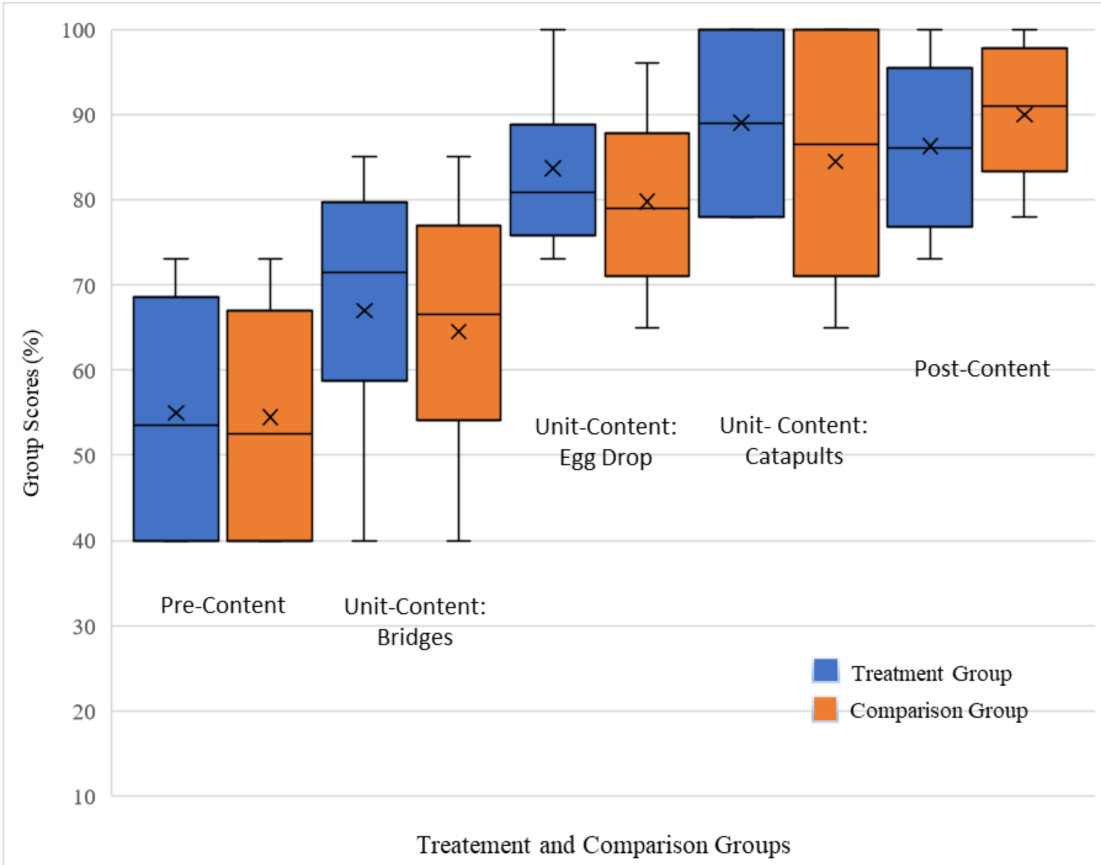


Figure 3. 6th grade P.E.T treatment group vs. comparison group scores of assessments, (*N*=20).

Overall, as depicted above, there is no statistical difference between the achievement of the treatment and comparison groups. A randomization test performed on the Pre-Content Assessment of each group yields a p-value of .96, thus demonstrating that if the scores across each group were mixed around in every combination possible, the outcome that was achieved is 96% likely; thus showing there is no difference in beginning achievement between the two groups. As the trimester went on and Unit Content Assessments along with the Post-Content Assessment were taken into consideration, independent t-tests were used to assess if the mean difference between the achievement of the two groups was statistically significant. p-values of .63, .65, .78, and

.48 were calculated through the t-tests. This shows that, as far as mean achievement of each class goes, the results were 63%, 65%, 78%, and 48% likely (all percentages are not statistically significant if an alpha value of 0.05 is used). This same lack of difference is also demonstrated when student answered Likert-Scale questions are analyzed with a Chi-Squared test of independence. In the end, the majority of students from each group agreed or strongly agreed that they were able to achieve a *Proficient* or *Advanced* (an B or an A) and that they were able to do this on their first try by the end of the trimester. The Chi-Squared p-values on the comparison of each group in relation to these questions shows that if there was no difference between the answers of the two groups, what was given was 98% likely and 67% likely (once again, percentages that are not statistically significant if an alpha value of 0.05 is used).

There is a likelihood that the over-all lack of difference between the treatment and comparison groups is based on the fact that the curriculum of the course was arranged in a spiral, or cyclical, manner (students attempted to achieve *Proficient* or *Advanced* on each standard multiple times over the course of the trimester). As students did this, they were given multiple opportunities to better be able to define engineering problems and what was needed for expected achievement. Therefore, as a whole, students showed better achievement by the end of the trimester.

Though there may have not been a significant difference overall, there appears to be a significant difference between the lower 50% of students from each group. If Figure 2 (shown above) is analyzed it can be seen that both groups have a similar spread of achievement during the first content unit. But, as the units progress, the lower 50%

students of the treatment group gain greater achievement in larger leaps. If assessments are delved into, both groups start off with struggling at identifying all criteria and constraints along with being able to accurately depict what the problem is. But, by the last content assessment, all but 1 of the treatment group students in the lower 50% were only 1 criteria identification away from a proficient (with the other student actually achieving *Proficiency* [a B]) while the students in the comparison group struggled with the identification of multiple criteria and/or constraints.

This marked difference may well be because of the self-assessment practice that the treatment students went through. If responses to Likert-Scale survey questions are analyzed, three of the four students from the treatment group that struggled the most shared similar beliefs to the following quote, “I get wut (sic) I need to do but I have a hard time getting ther (sic).” It may be likely that because these treatment group students knew what to do to achieve a *Proficient* or *Advanced*, but struggled at getting there, their participation in the self-assessment process helped them see what they were lacking in so that they could more intently focus on it the next time they defined an engineering problem.

Participation in self-assessment may have not only affected the lower 50% of students, but it also helped students who are typically “high achievers” gain greater achievement. This claim is not backed well with the numerical data from *Figure 2* and is also not detected with statistical data from the Likert-Scale survey questions that students completed. But, when specifically asked about the effect of self-assessment on their ability to achieve a *Proficient* or *Advanced* status when defining an engineering problem, the higher achieving students in the class who were in the treatment group claimed that self-assessment helped. These students elicited responses such as, “it helped me see where I was to see if I could make some improvements,” and, “it allows you to look back at yourself, kind of, saying ‘hey, I need to fix this.’” Self-assessment had a solid impact on some of my “higher achieving” students because it allowed them to go back over what they did to make sure they achieved what they were expected or wanted to. As one student put it, self-assessment helped them gain greater achievement, “because you can go back and read and see what you got when you self-assessed”.

Eighth Grade

The spread of data from the eighth grade treatment and comparison groups over the course of the trimester is shown below in Figure 3 through the use of box and whisker plots.

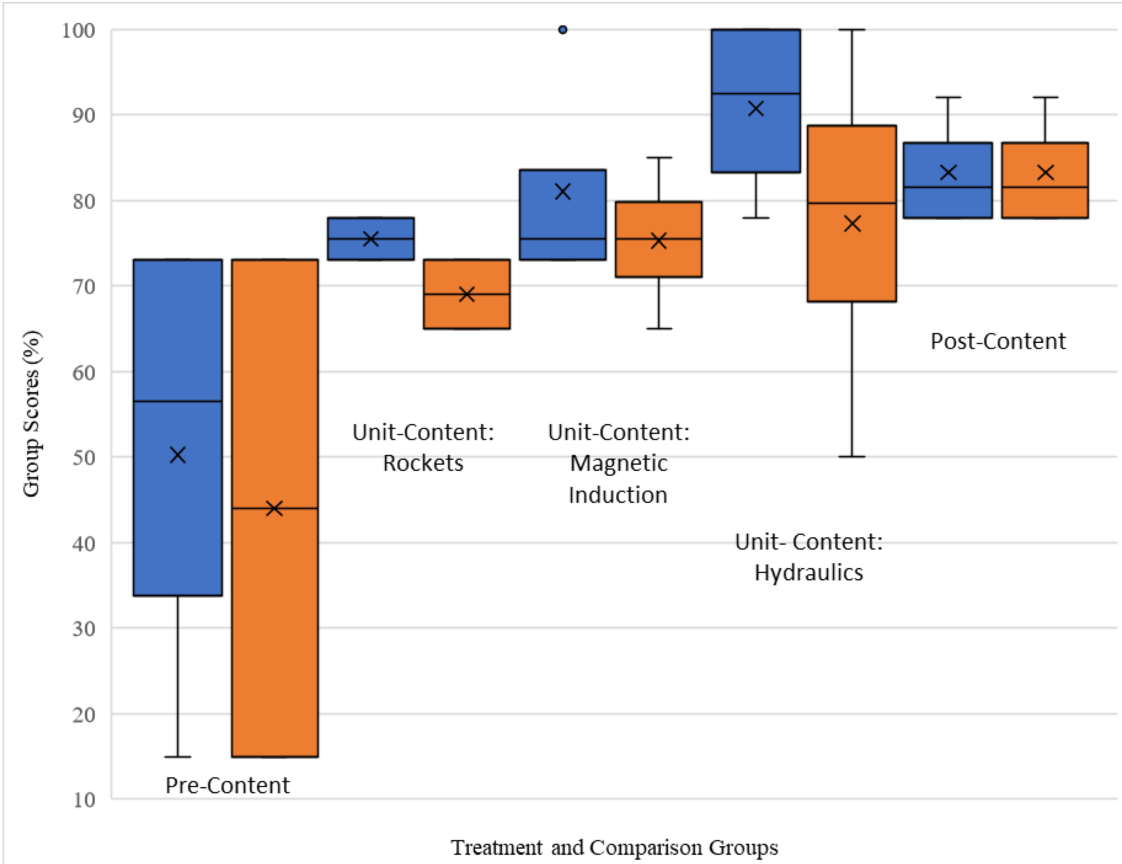


Figure 4. 8th grade P.E.T treatment group vs. comparison group scores of assessments, (*N*=14).

Based on the above figure, it can be seen that, starting out, both groups began with a similar spread in regard to understanding, with few members of each group being able to understand how to define an engineering problem in a *Basic* way (a C grade) and with the remainder of the students understanding how to define an engineering problem in a *Below Basic* way (an F grade). If the scores from each group were put through a randomization test to show that no matter the combination of scores across groups there is no statistical difference between the achievement of each group, a p-value of 0.87 would be obtained. This number expresses that if the scores from each group were mixed up in every way possible, the scores that were achieved are 87% likely. As this number is far from statistically significant, it can be accepted that there is no difference between the starting achievement of each group. They both started off at the same level of understanding because none of the students had defined engineering problems before.

A large jump can be seen from the Pre-Assessment to the first assessment, in both groups, but it can also be seen that there is a marked difference between the achievement of the treatment group and the comparison group. Most students in the comparison group achieved a *Below Basic*, which is a failing grade, while all students in the treatment group passed with *Basic* achievement (a C grade). As the unit assessments progress it can be noted that the treatment group continually achieved greater marks than those of the comparison group. If the mean differenced of both groups are placed into an independent t-test to find if there is a statistical difference between the achievement of the two groups, it can be found and there is strong evidence to suggest that self-assessment was the cause of this difference in achievement between the two groups. The following p-values were calculated from the t-tests (in order of assessment): 0.01, 0.12, 0.11. Each of these values suggest that if there was no difference between the mean achievement of each group, the specific scores that were obtained would happen 1%, 12% and 11% of the time. The first of the 3 values is considered statistically significant (it is below an alpha value of 0.05) and the other two are very close to being statistically significant. It would seem, that because the only difference between the two groups was that self-assessment was utilized among the treatment group and not among the comparison group, that self-assessment is a likely cause of the greater achievement among the treatment group.

It can be observed in Figure 3 above that the treatment group made a large jump in achievement between the second and third unit assessment while some members of the comparison group made considerable digression between the same assessments. During the Pre-Assessment and the first two units it was observed by the author that many students struggled with identifying and understanding the difference between engineering criteria and constraints. As such, a lesson was created and administered to all students in the class to help them understand these engineering principles. It would seem that this lesson, in conjunction with self-assessing, is also part of what helped the treatment group achieve a larger jump in their understanding as before the lesson, many were still struggling with achieving *Proficiency* (at least a B), and after the lesson, almost all achieved *Proficient* or *Advanced* (a B or an A).

As a last note, it must be communicated that this class of students was larger than 14. Because of the Covid-19 pandemic, 6 of the original 20 students in the class did not make it to school the first 3 to 6 days of class. As such, because of the uncertainty of their attendance, and their missing the Pre-Content Assessment, those students were not included in the study. If those students had participated in the study, there may have been a difference in the data gathered. The current data is accurate for those who participated in the study, but may not depict accurate results if the other students had taken part in the study.

In all, across all grade levels, there is reason to suggest that self-assessment did affect student achievement. Typically higher achieving students attribute some of their success to the practice of self-assessment, evidence of its effectiveness can even be seen in that of lower achieving students, and it seemed to make a large difference among the older group of students (for those who were able to take part in the study) as most of their mean scores were higher, in a statically significant (or close to it) way.

Effects of Self-Assessment on Confidence

In order to assess the effect that self-assessing had on student confidence when defining our engineering problems, a Likert-scale questionnaire was created to gauge each student's over all confidence when defining engineering problems. Data from these questions are shown below in Table 2.

Table 2. Likert-scale data of 6^{th} and 8^{th} grade student responses to confidence questions, (*N*=34).

Student Likert Scale Responses to the statement: By the end of the trimester, I was confident that I could achieve *Proficient* or *Advanced* on the engineering problems I defined.					
	1 Strongly Disagree	2 Disagree	3 Neutral	4 Agree	5 Strongly Agree
Treatment Group			4 (24%)	3 (18%)	10 (59%)
Comparison Group			1 (6%)	5 (31%)	10 (63%)
Student Likert Scale Responses to the statement: I am confident that if I was given another engineering problem to define, I could achieve a *Proficient* or *Advanced* on it.					
	1 Strongly Disagree	2 Disagree	3 Neutral	4 Agree	5 Strongly Agree
Treatment Group			4 (25%)	3 (19%)	9 (56%)
Comparison Group			3 (16%)	7 (37%)	9 (47%)

No students in either group, in either grade, disagreed with any of the confidence statements in the survey (shown above in Table 2), but all but one of the students who marked neutral seemed to have shown a lack of confidence if their written statements in regard to their answers are taken into consideration. When asked why they marked the answer they did, these students responded with answers like the following: “Because I know I will get at least one wrong”. This lack of confidence was likely because these students struggled with finding the criteria and constraints as is evidenced in the following quote from the survey, “I have a hard time listing criteria and constraints.”

Neutral answers aside, the data shows that most students felt confident in their ability to achieve *Proficient* or *Advanced* when defining their engineering problems, and this is all students from both groups. If the student responses shown in the above table for each statement are put through a Chi-Square Test of Independence to determine if any group strongly correlates with any response more than another group, the p-values come out as 0.69 and 0.83. Both numbers show that if one group did not correlate with a specific response more than the other, the results obtained would happen 69% and 83% of the time. As these are quite high percentages, the data from this survey shows that most students from both groups were equally confident in their abilities to achieve *Proficient* or *Advanced* on their engineering problems. This is likely because all students in both groups had much practice defining engineering problems throughout the trimester and all students became familiar with the process because of the cyclical nature of the curriculum. Evidence of this can be seen in Figure 2 as the spread of both group’s scores improve over the course of the trimester.

Another possible contributor to the confidence level of students in both groups may be that the author allowed students more than one attempt to achieve *Proficient* or *Advanced* (B's and A's) on the unit assessments. Because of this, many students, by the time they had taken the post-assessment, had around six or seven attempts at solving engineering problems and had become very familiar with the process of being able to define an engineering problem. As such, it is likely that their confidence grew along with their familiarity of the defining an engineering process.

Though it seemed the cyclical nature of the curriculum that the students went through is the cause of such high confidence among all students, when specifically asked about the role that self-assessment played in their confidence, the treatment group gave some insightful answers. Out of the nine students from the treatment group interviewed that were asked about the effect that self-assessment had on their confidence to achieve *Proficient* or *Advanced*, seven (78%) of the students replied that self-assessing boosted their confidence in some way. Student answers essentially reflected the following statement given by one of the students interviewed: "I can't see right off what I am doing right or wrong, and so it helped to be able to describe what I did…it just helped me visualize what I needed to do to get advanced." There seemed to not be much of a difference between the confidence of the two groups in being able to achieve *Proficient* or *Advanced*, but self-assessing did at least contribute to many students' confidence in being able to obtain high achievement.

Effect of Student Self-Assessment on Teaching

When data is looked at in regard to the effect that student self-assessment had on the authors teaching, a handful of themes emerged from the 12 journal entries written throughout the trimester the study was performed. These themes are laid out below in Table 3.

Table 3. Frequency of themes from the author's Teaching Journal.

Theme	**Frequency (Out of 12 Entries)**
Struggles and Successes with Identifying Engineering Criteria	9/12
Accuracy of student Self-Assessment	3/12
Lack of Understanding the Self-Assessment Rubric	2/12
Repetition Builds Confidence in Defining Engineering Problems	1/12

Of the 12 entries made by the author, 9 (75%) of them were regarding student struggles and successes with identifying specific engineering criteria when defining their problems. For the first while, students struggled with identifying both the criteria and constraints of their design problems. This was especially made evident by those who self-assessed. Many student self-assessments read similar to the following, "I identified all the criteria…I identified all the constraints." This was quite inaccurate as only two students identified all criteria and constraints during the first Unit-Content Assessment among both grades. This led to the realization that most of the students must struggle with understanding how to find criteria and constraints because many thought they were identifying them, when, in actuality, they were not.

In response, the author created a list of questions that the students could refer to that would help them better identify engineering criteria and constraints. This especially helped students identify constraints better as evidenced in this journal entry, “A lot more students are self-evaluating more accurately because of the questions related to the constraints.” More accurate self-assessment meant better achievement as it meant students were using the questions to help them work toward achieving *Proficient* (B) or *Advanced* (A) grades. After the first unit, most students no longer struggled with identifying engineering constraints, for the most part.

As the trimester progressed, students steadily became proficient at defining engineering problems, but many still struggled with identifying all necessary engineering criteria, and this showed in the accuracy of student self-assessments. This is evidenced by the following entry, “From these self-assessments, I am just really noticing that students are really struggling with identifying all the criteria.” This entry was made around the half-way mark of the trimester. It was later observed by the author that the, “students just need practice, especially in identifying criteria and constraints and the difference between them.”

About three weeks before the end of the trimester, an activity was created by the author and given to the students that scaffolded them through the process of identifying criteria and constraints of various design problems. After doing this, all students from both grades defined the engineering problem of the last unit, and, as an entry in the author's teaching journal states, "There was a marked difference in achievement in both groups," that resulted from the activity given to the students. This difference can be seen, particularly for the treatment groups, in *Figure 1* from earlier in this chapter. This activity happened between the second and third Unit-Content Assessments and a marked jump in achievement can be seen among both groups of students in both grades.

Then, as was alluded to toward the beginning of the chapter, the students participating in the study did not perform as was expected on their post-assessments on account of the students having a misconception as to what engineering criteria and constraints really are. The misconception is detailed as follows from the author's teaching journal, "In both grades there was a criteria where "Mearl" shouldn't have to lift her feet above 5 inches. Understandably, many students put this as a constraint, when, in reality, this was a criteria." Though this identification of a misconception was not directly the result of student self-assessment, it was included in this study because it was a major part of the study that helped the author identify a major misconception that students at that time, and likely students in the future, would have. Since then, the author has adjusted his teaching and curriculum to address the misconception and help students fully achieve *Proficient* or *Advanced* on the engineering problems they define.

Other notices from student self-assessment were noted by the author that impacted his teaching. One of these notices was that many students struggled at understanding the advanced portion of the rubric. As stated from the author's teaching journal, "My students don't understand my rubric…I am finding that, especially in regard to achieving advanced, student don't quite understand how it works." This was evidenced to the author because many students were self-assessing themselves at *Advanced*, but were not meeting the *Advanced* criteria. This was in the beginning of the trimester. As the trimester went on, and the students became more and more familiar with the rubric, along with the author learning how to better teach students how the rubric worked, more and more students started to understand the rubric better, particularly the advanced portion of it.

Another notice made by the author was that the repetition of the course seemed to build the confidence of the students. In his teaching journal, at a point that was around the half-way mark of the trimester (and by this time students had finished defining their second unit problem), the author wrote, "Students seemed more confident in their ability to define the problem at hand (this is not just those who formally self-assess)." After looking at student self-assessments and seeing that an increasing number of treatment group students were getting it, it was also seen that comparison group students were also understanding how to define engineering problems. By the end of the trimester close to 70% of students in each group, in each grade, were achieving *Proficient* or *Advanced* grades. As such, this confidence showed through among the self-assessors, particularly in the 6th grade, and self-assessments were becoming more and more accurate and in line with *Proficient* or *Advanced* standards (70% [7/10] of sixth grade student self-assessments were accurate and at least assessed at a proficient level; 57% [4/7] eighth grade student self-assessments were accurate and at least assessed at a proficient level). Even though self-assessment is not a direct measurement of confidence, it can be a byproduct of it, and this claim can be further solidified by the fact that at least 70% of students in both groups, in both classes, agreed that they were confident that they could get at least *Proficient* or *Advanced* when defining engineering problems by the end of the trimester and in the future.

In all, the author learned some key lessons from student self-assessment and from the results of the study itself to improve his teaching. First and foremost, the author learned that many students struggle with identifying criteria and constraints of engineering problems, and that is likely because many have never been exposed to performing that sort of practice. As such, it is important to equip students with questions to ask themselves that would help them identify the criteria and constraints of any engineering problem. It is also important to give the students plenty of practice in identifying engineering criteria and constraints as it is such a new concept, and this practice will eventually lead to greater achievement and greater student confidence. An additional major lesson learned, though not gleaned directly from student self-assessment, is that many students will likely have a misconception as to what engineering criteria and constraints really are and may get them confused. Through the identification of this misconception, the author will then be able to address this earlier with his incoming classes to allow students to better understand how to define engineering problems accurately.

CLAIM, EVIDENCE, AND REASONING

Claims From the Study

The main research question of this study was to address if student self-assessment affected student achievement when defining an engineering problem. Based on the data, and in answer to this question, self-assessment does influence student achievement when defining engineering problems. In most cases, when mean scores of treatment and comparison groups are compared, there is not a statistically significant difference between the scores that proves that self-assessment influenced greater achievement, though the scores do approach statistically significant in many cases (as indicated by the p-values of 0.32, 0.14, and 0.18 when T-Tests were conducted on student scores of Unit Content Assessments). If students are specifically asked about if self-assessment helped them have greater achievement when defining engineering problems, most students stated that it did help them gain greater achievement (78% of students interviewed reported this). The fact that self-assessment helped most students have better achievement is likely because the students who self-assessed had so much exposure to what was expected of them and they focused on it so much, that they were able to, eventually, better understand how to achieve *Proficient* or *Advanced* status when defining engineering problems. It is likely that this consequently also lead to students achieving better.

In addressing the second research question, which asked if self-assessment affected student confidence when defining engineering problems, data affirms that self-assessment does affect student confidence when defining engineering problems. Likert-Scale data taken from the study does not affirm this claim, but when students were specifically asked about it in interviews, 78% of students in the treatment group stated that self-assessing helped their confidence in being able to, at least, *Proficiently* define their engineering problems. It is likely that this is the case because of similar reasons found for the previous research question: because students were exposed to the criteria they were expected to meet so often and so much that they became more and more familiar with it and were then able to have better confidence in achieving better over time.

Regarding the third research question of this study, the analysis of student self-assessment greatly affected the author's teaching practice. As the author analyzed many student self-assessments he was able to identify that many students in his classes, not just those in the treatment group, were struggling with identifying engineering criteria and constraints (this was identified in nine out of 12 journal entries). Because the author was able to track student progress and specifically see student misconceptions among those who self-assessed (which also applied to those who did not self-assess), eventually all students were able to be helped to be more able to, at least *Proficiently*, define engineering problems.

Value of the Study and Consideration for Future Research

There is much value to this study, at least to the author, when it comes to teaching implications and student learning.

When it comes to influencing teaching practice, analysis of student self-assessment is greatly beneficial. In the study the author used student self-assessment as a type of formative assessment to see if students really understood what they were expected to understand. Doing this helped the author uncover student misconceptions and struggles regarding engineering, which then enabled him to create specific instruction to help clarify the mis-conceptions and struggles for his students. If the practice of analyzing student self-assessment is applied to other subjects and classes, it is likely that the same outcome could be had in those classes, thus helping a teacher boost student achievement.

This study greatly influenced student learning in a few ways. One of those ways is that students who participated in the self-assessment portion of the study were able to boost their achievement when defining engineering problems and their confidence at being able to at least *Proficiently* define engineering problems. It is likely that if this practice was introduced in other subjects that the students of those other subjects would have a similar reaction as the students in the present study.

Another way that the study influenced student learning is that it introduced students to the standards that they were expected to meet in a constant and dependable way. Because the students in the study, particularly those who self-assessed, were exposed repeatedly to the criteria that they were going to be measured by, which was presented to them in the same format every time, they knew what to expect by the end of the trimester and were better able to take on the challenges given to them.

One last way that this study influenced the students involved is that those who were in the treatment group learned a valuable soft skill that will aid them in whatever endeavors to participate in later in life. They not only learned how to accurately define an engineering problem at a potentially faster rate, but they also gained a skill that will aid them in the work-force later in life, particularly if they obtain a STEM (Science, Technology, Engineering, and Mathematics) job.

Though the study made an impact on the author's teaching practices and the learning of the students who participated in the study, future research on the topic of the study would be wise. As the study was only conducted over the period of a trimester (three months), the long-term effects of the present study are not known. Finding the long-term effects of the study would help to validate or refute the claims made in this study as whole class populations were not able to be used because of student absences due to the Covid-19 pandemic.

Investigating the effects of student self-assessment just on lower achieving groups may also turn out to be interesting. Research on this subject has already been performed, but most has not been done in an engineering context or in the context of a spiral curriculum, which was the context of this study. In this study it was found that self-assessment may have benefitted lower achieving students, which has not been the case in other studies, and it may be because of the use of self-assessment in a spiral curriculum.

Another potential area of study would be to study the effects of achievement over time without the spiral curriculum. In the present study it was seen that, particularly with the eighth grade class, the treatment group had greater achievement at a faster rate than the comparison group in the beginning of the study. Then, toward the end of the study, the comparison group started achieving about the same as the treatment group. It may be likely that the comparison group caught up to the treatment group because of the cyclical nature of the class curriculum. Perhaps there is a likelihood that if the curriculum was designed to go through topics and standards in a sequential manner (instead of a cyclical manner) that the treatment group would show increasingly greater achievement than that of the comparison group.

In the possible performance of future research, other designs could be used, but the design of the present study did turn out to be effective in finding how effective the treatment was. It was beneficial in such a way that there was a differentiation between the treatment group and the comparison group, which helped to find the difference in the performance of the two groups. As future research would likely be best based on a comparison between treatment and non-treatment groups to find the difference the treatment made, perhaps the present research design would be best to use to find future answers.

Impact of Action Research on the Author

This action research project and process had a large impact on the author in three main ways. The first of those ways is that this process influenced the author to formally see what research had already been done about the subject of self-assessment. In doing this the author learned valuable insights into his teaching practice. These insights included how self-assessment is not effective among students unless it is explicitly taught how to do it, self-assessment is best used as a formative assessment, and the more students self-assess, the better they become at it and the greater their achievement will be.

The action research process also helped the author set up his curriculum in a way to find out if a practice that he valued, self-assessment, was actually effective in his classroom or not. Previous to this process, the author questioned, and was not quite able to find out, if self-assessment was effective in his classroom. Knowledge of the action research process will allow the author to further find the effectiveness of other valued teaching practices in his classroom.

Lastly, the participation in the action research helped the author become a better professional teacher. Prior to participating in the process, as the author still doesn't yet even have five years of full-time teaching experience, the author was at times effective, but at other times ineffective, at identifying student misconceptions and struggles and addressing them. After performing this action research project, the author is better able to find valued student misconceptions and address them through focused and targeted interventions. Because of this process, the teacher will be able to be much more helpful to future students earlier in their acquisition of engineering learning.

REFERENCES CITED

Adeyemi, A. A. (2012). Effect of peer and self-assessment on male and female students' self-efficacy and self-autonomy in the learning of mathematics. *Gender & Behaviour, 10*(1), 4492-4508. https://search-proquest-com.proxybz.lib.montana.edu/docview/1021143055?accountid=28148

Almarode, J., Fisher, D., Frey, N., Hattie, J. (2018). *Visible learning for science, grades k-12: what works best to optimize student learning*. Corwin.

Andrade, H. L. (2019). A critical review of research on student self-assessment. *Frontiers in Education*, *4*(*87*). https://doi.org/10.3389/feduc.2019.00087

Bourke, R. (2016). Liberating the learner through self-assessment. *Cambridge Journal of Education, 46*(1), 97-111. http://dx.doi.org.proxybz.lib.montana.edu/10.1080/0305764X.2015.1015963

Elder, A. D. (2010). Children's self-assessment of their school work in elementary school. *Education 3-13, 38*(1), 5-11. http://dx.doi.org.proxybz.lib.montana.edu/10.1080/03004270802602044

Fastre, G. M. J, van der Klink, M. R., Sluijsmans, D., van Merrienboer, J. J. G. (2012). Drawing students' attention to relevant assessment criteria: effects on self-assessment skills and performance. *Journal of Vocational Education & Training*, *64*(*2*), 185-198. https://doi.org/10.1080/13636820.2011.630537

Gehringer, E. F. (2017, June), *Self-assessment to improve learning and evaluation.* Paper presented at 2017 ASEE Annual Conference & Exposition, Columbus, Ohio. https://peer.asee.org/28816

Harris, L. R., Brown, G. T. L. (2013). Opportunities and obstacles to consider when using peer- and self-assessment to improve student learning; case studies into teachers' implementation. *Teaching and Teacher Education*, *36*, 101-111. http://dx.doi.org/10.1016/j.tate.2013.07.008

Heritage, M. (2009). Using self-assessment to chart students' paths. *Middle School Journal*, *40*(*5*), 27-30. https://www.jstor.org/stable/23044491

Martin, C. S., Polly, D., & Kissel, B. (2017). Exploring the impact of written reflections on learning in the elementary mathematics classroom. *The Journal of Educational Research*, *110*(5), 538-553. http://dx.doi.org/10.1080/00220671.2016.1149793

McDonald, B. (2013). Mentoring and tutoring your students through self-assessment. *Innovations in Education and Teaching International, 50*(1), 62-71. https://doi.org/10.1080/14703297.2012.746516

McDonald, B., & Boud, D. (2003). The impact of self-assessment of achievement: The effects of self-assessment training on performance in external examinations. *Assessment in Education: Principles, Policy & Practice*, *10*(*2*), 209-220. https://doi.org/10.1080/0969594032000121289

NGSS Lead States. (2013). Next Generation Science Standards: For states, by states. The National Academies Press.

Ormrod, J. E. (2012). *Essentials of educational psychology: Big ideas to guide effective teaching*. Pearson

Semana, S., & Santos, L. (2018). Self-regulation capacity of middle school students in mathematics. *ZDM: The International Journal on Mathematics Education*, *50*(4), 743–755. http://dx.doi.org.proxybz.lib.montana.edu/10.1007/s11858-018-0954-0

APPENDICES

APPENDIX A

IRB EXEMPTION

INSTITUTIONAL REVIEW BOARD
For the Protection of Human Subjects
FWA 00000165

MONTANA STATE UNIVERSITY

2155 Analysis Drive
c/o Microbiology & Immunology
Montana State University
Bozeman, MT 59718
Telephone: 406-994-4706
FAX: 406-994-4303
E-mail: cherylj@montana.edu

Chair: Mark Quinn
406-994-4707
mquinn@montana.edu
Administrator:
Cheryl Johnson
406-994-4706
cherylj@montana.edu

MEMORANDUM

TO: Dylan Nelson and Walter Woolbaugh

FROM: Mark Quinn *Mark Quinn cj*
Chair, Institutional Review Board for the Protection of Human Subjects

DATE: September 1, 2020

RE: *"Effects of Student Self-Assessment on Achievement of an Essential Standard" [DN090120-EX]*

The above research, described in your submission of September 1, 2020, is exempt from the requirement of review by the Institutional Review Board in accordance with the Code of Federal regulations, Part 46, section 101. The specific paragraph which applies to your research is:

_X_ (b) (1) Research conducted in established or commonly accepted educational settings, involving normal educational practices such as (i) research on regular and special education instructional strategies, or (ii) research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods.

_X_ (b) (2) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, unless: (i) information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; and (ii) any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability, or be damaging to the subjects' financial standing, employability, or reputation; and (iii) the information obtained is recorded by the investigator in such a manner that the identity of the human subjects can readily be ascertained, directly or through identifiers linked to the subjects, and an IRB conducts a limited IRB review to make the determination required by section 16.111(a)(7).

____ (b) (3) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures, or observation of public behavior that is not exempt under paragraph (b)(2) of this section, if: (i) the human subjects are elected or appointed public officials or candidates for public office; or (ii) federal statute(s) without exception that the confidentiality of the personally identifiable information will be maintained throughout the research and thereafter.

____ (b) (4) Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available, or if the information is recorded by the investigator in such a manner that the subjects cannot be identified, directly or through identifiers linked to the subjects.

____ (b) (5) Research and demonstration projects, which are conducted by or subject to the approval of department or agency heads, and which are designed to study, evaluate, or otherwise examine: (i) public benefit or service programs; (ii) procedures for obtaining benefits or services under those programs; (iii) possible changes in or alternatives to those programs or procedures; or (iv) possible changes in methods or levels of payment for benefits or services under those programs.

____ (b) (6) Taste and food quality evaluation and consumer acceptance studies, (i) if wholesome foods without additives are consumed, or (ii) if a food is consumed that contains a food ingredient at or below the level and for a use found to be safe, or agricultural chemical or environmental contaminant at or below the level found to be safe, by the FDA, or approved by the EPA, or the Food Safety and Inspection Service of the USDA.

Although review by the Institutional Review Board is not required for the above research, the Committee will be glad to review it. If you wish a review and committee approval, please submit 3 copies of the usual application form and it will be processed by expedited review.

APPENDIX B

DEFINING ENGINEERING PROBLEMS GRADING RUBRIC

	4-Advanced	**3-Proficient**	**2-Basic**	**1-Below Basic**
GRADING CRITERIA	All Proficient Criteria are met AND the student goes above and beyond in a way that enhances their problem definition. Ideas for going above and beyond include, but are not limited to: • Stating an additional **3** more criteria that are **RELEVANT** to the problem. • Stating an additional 3 more constraints that are **RELEVANT** to the problem **The options provided are not the only options.**	The Problem Statement is no more than 2 sentences **AND** accurately describes the problem at hand. The Design Statement **clearly** shows that the solution will either be a device or a process. The Design Statement **clearly** explains what the device or process will do. The Criteria section includes the criteria set forth in the problem paper. The Constraints section includes at least the constraints set forth in the problem paper.	3-4 of the *Proficient* criteria have been met while 1 or 2 *Proficient* criteria have not been met.	1-2 of the *Proficient* criteria have been met while 3 or more *Proficient* criteria have not been met.

APPENDIX C

PRE-CONTENT/POST-CONTENT ASSESSMENT: PROMPT AND ACCOMPANYING DESIGN BRIEF

6^{th} Grade Pre/Post Test Prompt

The Predicament of Great Grandma Mearl

Great Grandma Mearl has lived in the same house since birth. Just like most Great Grandmothers, Mearl has countless memories in this house. Unfortunately, due to age, Mearl is no longer able to walk down the front stairs of this special house without assistance, not even with her cane. When Mearl needs to go somewhere that involves going down the front steps, Mearl needs help from children and grandchildren that are close. Because independence is important to Mearl, she wants to be able to leave this house without her children or grandchildren having to drop what they are doing to give her the assistance she needs. Because of this, Mearl has asked you to engineer a solution to her problem. This solution needs to be within a budget of $2,500 as this is all Mearl can afford. It also needs to not involve Mearl not having to lift her legs higher than 5 inches. Mearl desires that this solution be professional looking, as well as extremely durable as she hopes to use it the rest of her life.

6^{th} Grade Pre/Post Test Definition of Problem

Design Brief	
Client	Great Grandma Mearl
Designer	Your Name:
Problem Statement	Explain What the Problem Is:
Design Statement	I will design and test a process or device (circle one) that will:
Criteria	List the Criteria of the Problem:
Constraints	List the Constraints of the Problem:

8th Grade Pre/Post Test Prompt

The Predicament of Great Grandma Mearl

Great Grandma Mearl has lived in the same house since birth. Just like most Great Grandmothers, Mearl has countless memories in this house and cannot stand the thought of leaving her beloved home. Unfortunately, due to old age, Mearl is no longer able to walk down the front stairs of this special house without assistance, not even with her cane. For a while she was able to manage going down the stairs with one hand working the cane and the other hand leveraging her body with the railing, but she is no longer strong enough to pull herself up the stairs with the railing or lift her leg up each 10-inch step. When Mearl needs to go somewhere that involves going down the front steps, Mearl needs help from children and grandchildren that are close. Because independence is important to Mearl, she wants to be able to leave her house without her children or grandchildren having to drop what they are doing to give her the assistance she needs. Because of this, Mearl has asked you to engineer a solution to her problem. This solution needs to be within a budget of $2,500 as this is all Mearl can afford. It also needs to involve Mearl not having to lift her legs higher than 5 inches as well as not having to pull her body up more than 5 inches per every foot moved. Mearl desires that this solution be professional looking because she will have to see it every day. Lastly, Mearl wants the solution to be extremely durable as she hopes to use it for the rest of her life.

8th grade Pre/Post Definition of Problem

Design Brief	
Client	Great Grandma Mearl
Designer	Your Name:
Problem Statement	Explain What the Problem Is:
Design Statement	I will design and test a process or device (circle one) that will:
Criteria	List the Criteria of the Problem:
Constraints	List the Constraints of the Problem:

APPENDIX D

SIXTH GRADE UNIT CONTENT ASSESSMENT PROMPTS

Bridge Unit

The people of Microidia have a problem, a big problem! They live on an island and the only thing that connects their civilization with anything else is a bridge. Now, keep in mind that the people of Microidia are very, very small people. They are only about ½ of an inch tall. They depend on their bridge to get off their island so that they can get food and other supplies needed to live. The problem is that their bridge is so old and worn out that it is close to breaking. That is why you are here! The people of Microidia have asked that a panel of engineers (all of you) come up with different design solutions that meet their needs. From the different design solutions that you come up with, after testing, one design will be submitted to the Microidians for their final bridge.

The Microidians are allergic to every building material besides wood, yarn, and Elmer's Glue. Because the bridge is collapsing soon, a final bridge needs to be built and ready to test in 8 class periods. As bridges can become quite expensive, the Microidians expect you to spend no more than $16,000 tops.

The bridge that you build needs to be 30 cm. in length and must be tall enough for a Microidian to comfortably walk through. It also must have 2 bases that are 5cm. X 5cm. It is also important to note that there should be nothing in the middle of the bridge that is touching the ground. Apart from that, the sky is the limit on how you build your bridge. It is up to you and your partner to research bridge designs and come up with a design that both of you feel is best to do the job.

In the end, the Microidians are going to want the bridge that costs a reasonable amount, is strong for its cost, and can be built in a reasonable amount of time.

Good luck and work hard!

Egg Drop Unit

Terrorists have imported nuclear eggs into the US and they have all landed in Rigby, ID. National experts have figured out a way to un-detonate the eggs: by dropping them from the a tall height in such a way that when they hit the ground the eggs won't break. Doing this releases a chemical reaction to the radioactive material inside the egg that makes it non-atomic. Unfortunately, these national experts were assassinated, and the responsibility has been turned to you. The Terrorists wouldn't suspect middle schoolers to un-detonate all of those eggs.

To accomplish this task that has been bequeathed to you, it is expected that you build a contraption to go around an atomic egg that will make it so that the egg doesn't break when it hits the ground. Additionally, it is expected that this contraption utilizes a parachute. Because quality work is expected, your contraption is expected to be professional looking. For a successful contraption it is also expected that at least 6 class periods are used to build and refine your contraption to be its best.

In this scenario, time truly is of the essence. You have a total of 8 days to design and test your contraption. Furthermore, the following supplies are what you have to work with: Rubber bands, yarn, paper, 10 in. tape, garbage bags, straws, cardboard substances. If you have an idea of another substance that you would like to use, run it be Mr. Nelson to be cleared to use.

Bear in mind that if just one egg breaks the entire world will explode.

Work hard and work smart. Human life as we know it is on the line!

The Nation thanks you for your service and wishes you good luck!

Catapults

Hear ye, hear ye! War has been declared!

The *Imperpetuum Stone* has been rumoured to have existed for thousands of years, and only just recently have the coordinates of it been discovered! Obtaining the *Imperpetuum Stone* would mean that your country would gain a stone with the power to grace all living in your country and all the family that comes after you with shear awesomeness and womanly/manly brawn and power. Every country (table) in the world (the classroom) wants this stone, therefore, your local government has come to you to ask you to build a catapult to destroy the other govermnents in the world so that your government can obtain the *Imperpetuum Stone*.

It is expected that this catapult that you build be extremely well built as it would be bad if it fell apart while you were conquesitng to take over the world for the *Imperpetuum Stone*. As well, because you are at war with the other countries (tables) in the world (the classroom) your catapult is expected to be able to launch objects to the country that is nearest to you and to the country that is farthest from you.

It is to take no more than 4 class periods to build this catapult as time truly is of the essence. The materials you are expected to use are popsicle sticks, yarn, rubber bands, plastic spoons. You are welcome to bring in other materials if it is cleared by Mr. Nelson. The supplies that you use is to cost no more than $20.

REMEMBER: FIGHT FOR YOUR COUNTRY VALLIANTLY! BUILD A CATAPULT WORTH OF HONOR AND PRAISE. DON’T LET YOUR COUNTRY DOWN!

APPENDIX E

EIGHTH GRADE UNIT CONTENT ASSESSMENT PROMPTS

Rockets

EARTH: Exit Premises

Planet Earth has issues! The planet has been over-run by zombies, global warming is rapidly decreasing food and fresh water, and many natural resources necessary for human survival are depleting. The human race is in grave danger.

Astronomists have located a distant planet within the milky-way galaxy that just might be our saving grace. This planet has been named ***PLANET X***. It is un-inhabited but is extremely similar to Earth. The global temperature is stable and there is plenty of natural resources to support humanity.

As the fate of the world is quickly coming to an end, astrophysicists and engineers are turning to you to solve our earthly problems and get us off of this planet and onto ***PLANET X*** by creating the best possible rocket.

As ***PLANET X*** is a very long ways away, it is expected that you create a rocket that will go the farthest distance possible. Because many people will be riding in the rocket, it is expected that it look very professional. As well, the expectations stands that it should be built strong enough so that it does not explode at take-off.

Because it is almost the end of the world resources are scarce. For each rocket you are only to use an 11 X 8.5 sheet of material and 10 inches of tape. As time truly is of the essence your rockets must be built and tested by 6 days (6 class periods). As materials cost money, it is expected that each rocket stay below the cost of $100 Million (each square centimetre of material costs $165,775).

Magnetic Induction

These days are dark days in the little town of Rigby, Idaho…dark days I tell you…no, like, literally dark, as in there is absolutely no light whatsoever. Fossil fuel resources are becoming more and more scarce and because of such, electrical power is no longer able to be generated for the town and surrounding parts of Rigby. The town simply will not manage without electrical power. That is what runs our refrigerators, freezers, lights, heat, some of our cars, and this is also what gives us electricity to generate light and use our devices.

Because of this major problem a panel of designers (all of you) have been selected to design a model renewable energy solution to this problem. It is expected that this solution utilizes magnetic induction. It is expected that the renewable energy design you create will be compatible with the environment and resources unique to the Rigby area. This model that you generate is expected to generate at least .3Volts of electricity.

As this is an urgent matter, time is very much of the essence. You have 6 class periods to design, build and test this product. You are also limited on the amount of cardboard you can use. It is expected that you use no more than 1000 square centimetres of cardboard, with each square centimetre of cardboard costing $2,000. Your budget is not to exceed $2 Million. The supplies you are allowed to use is the following: 1 spool of magnet wire, 14 rare-earth magnets, cardboard, wire for a shaft, and hot glue (you are welcome to bring in your own magnet wire and rare-earth magnets). If you have other materials that you would like to use, run it by Mr. Nelson and he will let you know if it is okay to use.

Your community is depending on you. Work hard and work smart!

Hydraulics

Here it is, the long-awaited hydraulics project. You have now gone through the full engineering process a few times and now is the time to put your knowledge of how to solve an engineering problem to the test.

For this project you are to pick your own life problem or world problem to solve using a hydraulic machine. The life problem can be anything from sleeping through your alarm-clock to having a machine that will feed your pets for you. The world problem can be anything from using a hydraulic machine to clean the plastic out of the ocean to helping amputees to adapt the world around them. Really, the problem could be just about anything, but you must choose to solve EITHER a notable world problem or a problem you have in your life, and the problem must be able to be solved with a hydraulic machine.

Even though you are solving a problem that you pick, certain things are expected of your project. Whatever the contraption is that you build, it is expected that it uses at least 2 of the 3 classes of levers. It is also expected that this project be very well built.

You will have 5 class periods to build and test this contraption. The supplies that you have to use for this project are the following: hot glue, skewers, zip-ties, air-line tubing, skewers, cardboard. If you would like to use other supplies than those listed, clear it with Mr. Nelson before you bring it in to use. You are to spend no more than $25 total on the supplies on this project.

This is your opportunity to really make a project that is meaningful to you. Of all the projects done this trimester, this project will be the one that is easiest to reach advanced when defining the engineering problem. Work hard and work smart!

APPENDIX F

SELF-ASSESSMENT FORM

Defining the Problem: Student Self-Assessment Sheet

Participation in this research is voluntary and participation or non-participation will not affect a student's grades or class standing in any way.

<table>
<tr><th></th><th>4-Advanced</th><th>3-Proficient</th><th>2-Basic</th><th>1-Below Basic</th></tr>
<tr><td>GRADING CRITERIA</td><td>All Proficient Criteria are met <u>AND</u> the student goes above and beyond in a way that enhances their problem definition.

Ideas for going above and beyond include, but are not limited to:
Stating an additional 3 more criteria that are RELEVANT to the problem.
Stating an additional 3 more constraints that are RELEVANT to the problem
*The options provided are not the only options. If the student has felt they have reached advanced in some other way, they will need to describe it below. *</td><td>The Problem Statement is no more than 2 sentences <u>AND</u> accurately describes the problem at hand.

The Design Statement <u>clearly</u> shows that the solution will either be a device or a process.

The Design Statement <u>clearly</u> explains what the device or process will do.

The Criteria section includes the criteria set forth in the problem paper.

The Constraints section includes the constraints set forth in the problem paper.</td><td>3-4 of the Proficient criteria have been met while 1 or 2 Proficient criteria have not been met.</td><td>1-2 of the Proficient criteria have been met while 3 or more Proficient criteria have not been met.</td></tr>
<tr><td>STUDENT SELF-ASSESSMENT: Below the score you think you have achieved; explain why you feel you have achieved that score.</td><td></td><td></td><td></td><td></td></tr>
</table>

APPENDIX G

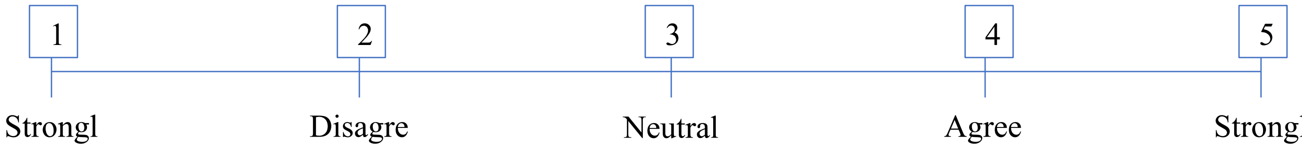
LIKERT SCALE SURVEY

Student Perceptions of Self-Assessment

Participation in this research is voluntary and participation or non-participation will not affect a student's grades or class standing in any way.

The purpose of this questionnaire is for Mr. Nelson to understand what your feelings and perceptions were about various activities that you have done this trimester. This will not be graded and will not affect your standing in this class.

Use the following graphic to help you answer the questions in this section:



1. I clearly understood what I needed to do to achieve *Proficient* or *Advanced* on the engineering problems that we defined.

1	2	3	4	5

2. I was able to achieve *Proficient* or *Advanced* on the engineering problems that we defined.

1	2	3	4	5

   a. In the space provided, please explain why you answered the way you did in the above question:

3. By the end of the trimester I was able to achieve *Proficient* or *Advanced* on the engineering problems we defined **on my first try**.

1	2	3	4	5

4. By the end of the trimester, I was confident that I could achieve a *Proficient* or *Advanced* on the engineering problems that I defined.

1	2	3	4	5

   a. In the space provided, please explain why you answered the way you did in the above question:

5. I am confident that if I was given another engineering problem to define that I could achieve a *Proficient* or *Advanced* on it.

1	2	3	4	5

6. I can accurately self-assess myself when defining an engineering problem.

1	2	3	4	5

‘APPENDIX H

STUDENT INTERVIEW QUESTIONS

Student Interview Questions

Before interviewing students, this phrase will be read: "**Participation in this research is voluntary and participation or non-participation will not affect your grades or class standing in any way**."

1. Did you participate in formal self-assessment during this trimester [show student the self-assessment sheet—to clarify, may need to ask if the student filled this out during the trimester]?

    a. Do you feel this affected how likely you were to achieve a *Proficient* or *Advanced* when we defined our engineering problems?

        i. Why? Please explain.

2. How confident were you in your ability to achieve a *Proficient* or *Advanced* when we defined our engineering problems?

    a. Did the fact that you did/didn't formally self-assess yourself affect this?
        i. How so?

The following questions are for students who self-assessed only:

3. By the end of the trimester, how accurate do you feel you were in your self-assessments when we defined our engineering problems?

    a. Can you explain why you feel this way?

        i. Can you give an example of when you were able to accurately self-assess yourself on how well you defined an engineering problem [have the self-assessment sheet handy to help them job their memory]?