The impact of confidence-based marking on unit exam achievement in a high school physical science course

Casey Clark

University of Northern Iowa

Copyright ©2020 Casey Clark

Follow this and additional works at: https://scholarworks.uni.edu/grp

Part of the Educational Assessment, Evaluation, and Research Commons, and the Science and Mathematics Education Commons

Let us know how access to this document benefits you

Recommended Citation

Clark, Casey, "The impact of confidence-based marking on unit exam achievement in a high school physical science course" (2020). Graduate Research Papers. 1449.

https://scholarworks.uni.edu/grp/1449

This Open Access Graduate Research Paper is brought to you for free and open access by the Student Work at UNI ScholarWorks. It has been accepted for inclusion in Graduate Research Papers by an authorized administrator of UNI ScholarWorks. For more information, please contact scholarworks@uni.edu.
The impact of confidence-based marking on unit exam achievement in a high school physical science course

Abstract
High school teachers begin each year with a plethora of content standards to teach, however, instructing students how to learn is hardly ever considered. Metacognition and learning how to learn are not present in high school science content standards. The ability to self-regulate our knowledge is essential to assumptions about learning (Cubukcu, 2009). Students are expected to use metacognitive strategies, even if they do not have them. When students work through problems, it is common to expect them to go back and recheck their work. This is a form of metacognition and self-regulation. Asking students apply metacognition can help them learn both content knowledge and gain metacognitive skills (Schwartz et al., 2009). From professional experience, students work systemically through practice problems and assessments by treating each problem the same. High achieving students are expected to show confident judgment in their knowledge, yet explicit feedback on quality of the knowledge is often not given by teachers. How to improve judgment in knowledge is taught even less often. Students are unable to express what knowledge they already possess. They also struggle to articulate to what degree they believe this knowledge to be accurate (Gahan & GardnerMedwin, 2003; Gardner-Medwin, 1995).

Measuring a student’s confidence in their knowledge can be a difficult challenge for educators. Confidence-based marking (CBM) is one approach to improve formative self-assessment and to help train students to become more objective of their answers when undergoing formative and summative evaluation. CBM is a grading scheme, commonly used for multiple choice and true or false questions, which requires the students to assess the selected answer and place a descriptor on it indicating how confident they are the answer is correct.

With a tool of this caliber at the aid of educators, a puzzle remains. Why is a well-researched, sensible, and practical strategy for objectively marked tests not employed by more teachers (Davies, 2002; Gardner-Medwin, 2006; Hassmén & Hunt, 1994)? The purpose of this study was to investigate the effectiveness of CBM in a high school physical science course. Research appears in medical colleges and similar post-secondary settings, primarily in preparation for the medical field, however, there is no clear published literature about implementation in the high school setting. There also exists ample research to show the effectiveness of the tool, but a lack of research to show if the age of the student influences its effectiveness. This study sought to determine if CBM can be effective at the secondary level in a science classroom.

This open access graduate research paper is available at UNI ScholarWorks: https://scholarworks.uni.edu/grp/1449
The Impact of Confidence-Based Marking on Unit Exam Achievement in a High School Physical Science Course

Casey Clark

University of Northern Iowa
This Paper by: Casey J Clark

Entitled: The Impact of Confidence-Based Marking on Unit Exam Achievement in a High School Physical Science Course

has been approved as meeting the non-thesis requirement for the

Degree of Master of Arts

Date: 7/23/2020   Dr. Kyle Gray, Advisor

Date: 7/23/2020   Dr. Jeff Morgan, Outside Reader
Acknowledgements

I would like to express my deepest appreciation to my advisor, Dr. Kyle Gray, for the assistance, support, and suggestion throughout my project. His positive attitude and endless knowledge of research helped guide me to success during my program. Without his guidance, teaching, and reviewing this project would not have been possible.

I would also like to thank my outsider reader, Dr. Jeff Morgan. His support and genuine interest in the project helped me continue to ask more of my paper and my project. In addition, I want to thank all the graduate staff in the Science Education Department at the University of Northern Iowa for commitment to enhancing my understanding of both teaching and myself while I made my way through the program.

In addition to all the staff at the University of Northern Iowa, I would like to thank the Urbandale Community School District. As I work there as an educator, they graciously allowed me to conduct my research in my classroom. Without their support of my dream of furthering my education this project would not have been possible. I would like to add an additional thanks to my immediate coworker, Jordan Sonntag, who through the planning, implementation, and writing of this project was there as a support and sounding board.

As a final, but certainly not last, acknowledgment I want to thank my wife, Amy Clark. During this program and project, she always provided me unwavering support and inspiration to persevere toward achieving my degree. She did so while giving birth to and raising our two beautiful daughters, Dani, and Bailey. Without all their support and encouragement, I would not have been able to continue to persist and achieve the goals I set out when starting this program.
# Table of Contents

Chapter 1: Introduction ............................................................................................................. 1

Chapter 2: Literature Review ................................................................................................. 3

Confidence-Based Marking ..................................................................................................... 3

Confidence-Based Marking Structure ..................................................................................... 4

Confidence-Based Marking Assessment Construction ............................................................ 7

Confidence-Based Marking Assessment .................................................................................. 10

Academic Benefits .................................................................................................................. 11

Concerns against Confidence-Based Marking ....................................................................... 14

Learning Management Systems ............................................................................................. 15

Standards Referenced Grading ............................................................................................... 16

Theoretical Framework ........................................................................................................... 17

Chapter 3: Methods .................................................................................................................. 20

Research Questions ............................................................................................................... 20

Participants ............................................................................................................................ 20

Site ........................................................................................................................................ 21

Data Collection Procedures ................................................................................................ 23

Researcher Role ..................................................................................................................... 25

Instructional Method Comparative Research ...................................................................... 26
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Materials</td>
<td>27</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>27</td>
</tr>
<tr>
<td>Chapter 4: Results and Discussion</td>
<td>31</td>
</tr>
<tr>
<td>Participants</td>
<td>31</td>
</tr>
<tr>
<td>Data Removal in Fall of 2019</td>
<td>31</td>
</tr>
<tr>
<td>Research Question 1</td>
<td>33</td>
</tr>
<tr>
<td>Learning target 1</td>
<td>33</td>
</tr>
<tr>
<td>Learning target 2</td>
<td>33</td>
</tr>
<tr>
<td>Research Question 2</td>
<td>34</td>
</tr>
<tr>
<td>Learning target 1</td>
<td>34</td>
</tr>
<tr>
<td>Learning target 2</td>
<td>36</td>
</tr>
<tr>
<td>Discussion</td>
<td>39</td>
</tr>
<tr>
<td>Research Question 1 Unit Exam Scores</td>
<td>39</td>
</tr>
<tr>
<td>Research Question 2 Practice Tests and Unit Exam Scores</td>
<td>41</td>
</tr>
<tr>
<td>Using CBM in the High School Classroom</td>
<td>43</td>
</tr>
<tr>
<td>Limitations</td>
<td>46</td>
</tr>
<tr>
<td>Further Questions and Continued Research</td>
<td>48</td>
</tr>
<tr>
<td>Conclusion</td>
<td>49</td>
</tr>
<tr>
<td>References</td>
<td>52</td>
</tr>
<tr>
<td>Appendix A: Learning Target 1 – CBM Practice Test and Unit Exam Samples</td>
<td>58</td>
</tr>
<tr>
<td>CBM Practice Test – Learning Target 1</td>
<td>58</td>
</tr>
<tr>
<td>Unit Exam – Learning Target 1</td>
<td>63</td>
</tr>
<tr>
<td>Appendix B: Learning Target 2 – CBM Practice Test and Unit Exam Samples</td>
<td>66</td>
</tr>
<tr>
<td>CBM Practice Test – Learning Target 2</td>
<td>66</td>
</tr>
<tr>
<td>Unit Exam – Learning Target 2</td>
<td>69</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

High school teachers begin each year with a plethora of content standards to teach, however, instructing students how to learn is hardly ever considered. Metacognition and learning how to learn are not present in high school science content standards. The ability to self-regulate our knowledge is essential to assumptions about learning (Cubukcu, 2009). Students are expected to use metacognitive strategies, even if they do not have them. When students work through problems, it is common to expect them to go back and recheck their work. This is a form of metacognition and self-regulation. Asking students apply metacognition can help them learn both content knowledge and gain metacognitive skills (Schwartz et al., 2009). From professional experience, students work systemically through practice problems and assessments by treating each problem the same. High achieving students are expected to show confident judgment in their knowledge, yet explicit feedback on quality of the knowledge is often not given by teachers. How to improve judgment in knowledge is taught even less often. Students are unable to express what knowledge they already possess. They also struggle to articulate to what degree they believe this knowledge to be accurate (Gahan & Gardner-Medwin, 2003; Gardner-Medwin, 1995).

Measuring a student’s confidence in their knowledge can be a difficult challenge for educators. Confidence-based marking (CBM) is one approach to improve formative self-assessment and to help train students to become more objective of their answers when undergoing formative and summative evaluation. CBM is a grading scheme, commonly used for multiple choice and true or false questions, which requires the students to assess the selected answer and place a descriptor on it indicating how confident they are the answer is correct.
Moving beyond traditional multiple-choice assessment is essential. The question type only provides data of right or wrong, leaving out information about student self-assessment or thought processes (Davies, 2001). CBM can aid in the breakdown of misconceptions by allowing students to assess their knowledge and judge how well they know and understand a topic. This extends to teachers as well as it can provide insight into student understanding.

With a tool of this caliber at the aid of educators, a puzzle remains. Why is a well-researched, sensible, and practical strategy for objectively marked tests not employed by more teachers (Davies, 2002; Gardner-Medwin, 2006; Hassmén & Hunt, 1994)? The purpose of this study was to investigate the effectiveness of CBM in a high school physical science course. Research appears in medical colleges and similar post-secondary settings, primarily in preparation for the medical field, however, there is no clear published literature about implementation in the high school setting. There also exists ample research to show the effectiveness of the tool, but a lack of research to show if the age of the student influences its effectiveness. This study sought to determine if CBM can be effective at the secondary level in a science classroom.
Chapter 2: Literature Review

Confidence-Based Marking

Providing confidence in knowledge or statements can be traced back to 1906 and Australian weather forecasting. These exercises used a five-point system to communicate how confident the forecaster was with their predictions of the upcoming weather (Cooke, 1906). In education, confidence-based marking (CBM) studies date back to the 1930s. The majority of the research is related to post-secondary education with research coming from medical and psychology programs. The University College London has used CBM since 2001 for summative medical exams for students in their first and second year (Gardner-Medwin, 2006). CBM demands a deeper level of understanding of content, confidence in knowledge while placing a premium on careful thinking and improved reliability. (Gardner-Medwin, 2006; Gardner-Medwin & Curtin, 2007; Hevner, 1932). Multiple-choice tests of knowledge are currently the most common form of self-assessment because they adequately sample the content and have a high level of reliability as well as being easy to score (Khan et al., 2001). In the medical field, it is imperative for the individual’s knowledge to be sound and be able to judge what they do not know. Through this grading scheme, it is possible to measure a student’s belief in their answer on an assessment, ultimately revealing their actual level of understanding of the topic. Student performance is ultimately affected by the knowledge they have stored in memory, and by how quickly and accurately it can be retrieved and processed (Hunt, 1982). CBM ultimately provides the educator insight into each student’s ability to retrieve, process, and employ this knowledge.

Students are able to improve their metacognitive accuracy over time as measured by exams. Multiple opportunities to practice and undergo metacognitive experiences may be the best approach to long term improvement (Callender et al., 2016). A bulk of the research into
CBM occurred before the computer-aided assessment was practical on a large scale (Ahlgren, 1969; Gardner-Medwin, 2006). However, the core of the practice has remained the same since the beginning, providing opportunities to improve assessments and self-reflection in students.

**Confidence-Based Marking Structure**

The two governing factors of CBM are confidence an answer will be correct, and the rewards or penalties for right or wrong answers (Gardner-Medwin, 2006). CBM commonly defines confidence as the degree of belief or the subjective probability a selected answer is correct. Research studies outline the widely used grading scheme for questions. Students answer questions presented to them, followed by a subquestion where they select their confidence in the answer. Confidence defined commonly as "C" has associated levels based on the level of confidence. This tiered three-point confidence scale in shorthand is C=1, 2 or 3. When an answer is correct with a chosen confidence, the score for the question is equal to its associated confidence level (score = 1, 2, or 3). If the selected answer is incorrect, then penalties are incurred with increasing harshness for misused higher confidence levels (Bloxham & Boyd, 2007; Gardner-Medwin, 1995, 2006; Hassmén & Hunt, 1994, p. 199).

For typical CBM assessments, students answer traditional multiple-choice or true/false questions. Upon completion of each question, students then select their level of confidence in the correctness of their answers. For a simple system, this would be selecting either low, medium, or high for the question. The answer is graded for correctness upon submission, and then the points earned for the question would be adjusted to reflect the level of confidence shown (Barr & Burke, 2013).
Alternatively, a descriptor may be used. It can be a phrase or number indicating a range of levels of confidence. A typical example of a descriptor is "I am very confident" or a number scale of one through three. An example of this is shown in Figure 1. Educators use this tool to measure students' beliefs in their knowledge accurately. Knowledge is not just a lucky guess, and a misconception is worse than acknowledged ignorance. The need to overcome misconceptions is essential for secondary science classrooms (Gardner-Medwin, 2006).

**Figure 1**

*Sample Confidence Based Marking Question*

In an example, a student correctly answers a question and selects high confidence. The score for this question is three points, equal to the confidence level selected. In contrast, a student answering the same question incorrectly and selects high confidence would incur a larger penalty than the highest possible positive outcome, negative six points. The points returned depends on the combination of answer correctness and confidence chosen. In a final example a student could earn one point for a question if they correctly answered but chose low confidence. The possible weight of negative points has changed throughout the history of CBM.
use, the typical negative value for high confidence on an incorrect answer is negative six points. From professional experience, secondary students experimenting with CBM for the first time often earn negative scores at the beginning, before they completely understand the scoring. Large negative scores lead directly to the purpose of using CBM: identification of lack of confidence or misconception (Gardner-Medwin, 1995, 2006; Hassmén & Hunt, 1994).

If students incorrectly assess their level of confidence, they demonstrate their lack of understanding of the assessed content. A well-designed CBM scheme ensures students receive the best scores when they differentiate responses based on content understanding. Having a clear understanding of the subject matter will earn higher scores on a confidence-based assessment. Conversely the option of selecting low confidence is also very beneficial. Low confidence benefits the student when there is a reservation because the penalties are proportionately less or absent when correctly judging one's confidence (Gardner-Medwin, 2006). If the student shows metacognitive awareness and chooses low confidence \((C = 1)\) for what they believe is an incorrect answer they would not receive negative points. By doing so they have correctly indicated the lack of knowledge of the assessed content.

Confidence-based marking (CBM), no matter the form, is not widely used outside of the post-secondary setting. In medical education self-assessment in the form of CBM is used to improve student's ability to judge their knowledge. It is a valuable gauge for assisting learners to become aware of their strengths and weaknesses (Khan et al., 2001). Under this framework, CBM shows connections to Albert Bandura's model of self-efficacy. Self-efficacy strength is increased through tasks which exercise expectations of personal efficacy. This opportunity determines whether or not coping behavior starts, the amount of effort expended, and how
long the individual will sustain it. Undergoing experiences that appear threatening but are safe procedures on the road to mastery continue to enhance self-efficacy and regulation (Bandura, 1977). This idea begins to connect CBM, a style of assessment question, and modes of learning.

Due to CBM's narrow scope of use, there is a lack of readily accessible information about CBM. Published materials exist about CBM and the uses however, anecdotally K-12 education teachers are not familiar with CBM, how it works, or why it could be a useful resource for assessment in their content. At University College London, students report CBM as simple, fair, readily understood, and beneficial. Completing CBM tasks test their knowledge with reward, while also allowing for options when there is a lack of confidence. In a 1998 survey of students, 95% of respondents replied confidence-based assessment was either useful or very useful in their learning (Gardner-Medwin, 2006; Issroff & Gardner-Medwin, 1998).

**Confidence-Based Marking Assessment Construction**

There is no clear consensus as to the best format when building a confidence-based marking (CBM) assessment. How confidence scales are displayed varies based on the source material as well as the scale and wording used for the confidence assessment task. Descriptors of confidence can vary significantly in both description and number of levels of confidence (Davies, 2002; Khan et al., 2001). These variations could be as significant as using terms to show confidence (i.e., very sure; fairly sure; neutral; unsure; pure guess) or a numbered scale (i.e., $C = 1, 2, 3$). Typically, CBM assessments utilize three to five levels of confidence. Regardless of which method, the commonality is the importance of a transparent scheme defined by rewards, penalties, and well-defined risks and rewards (Gardner-Medwin, 2006).
Early research utilized descriptors to help guide the reflective nature of the participants (Soderquist, 1936). At University College London confidence levels are deliberately identified by numbers or neutral descriptors (low, med, high) rather than by descriptive terms ('certain,' 'very sure,' 'unsure,' 'guess') because descriptors mean different things to diverse people in various contexts. This variation and uncertainty make CBM subjective, not objective (Echternacht, 1972; Gardner-Medwin, 2006).

True-false assessments at the University of Minnesota in 1935 (Soderquist, 1936) is an example of early CBM assessments. A 75 true-false item test was constructed requiring students to mark their degree of assurance on their responses. Soderquist used a positive point system with a doubled negative for incorrect confident answers. During the assessment, the system utilized allowed assessment takers to select the credit they wished to earn with scores ranging from one to four. For incorrect answers with a higher score (confidence) request, the penalty would be double. Adding confidence tasks to each question also had what some consider a drawback. The time added to the total assessment increased significantly (Soderquist, 1936).

Davies (2002) found the average time taken per student in answering CBM questions increases considerably. There are also indicators showing less guessing occurs under the methods of weighting responses for assurance than under the conventional assessment methods (Echternacht, 1972; Soderquist, 1936). Presumably, as there were more substantial penalties for high confidence, incorrect answers, students became more conservative in their responses. As students have increased conservatism, it results in being driven to be more
assured with answers. Students claim CBM assessments are harder to take and they ‘have to think’ when performing the confidence tests (Davies, 2002).

Automated assessments in the form of multiple-choice and true-false questions are beneficial to educators as they are time reducing in grading. These test types come with increased efficiency and the removal of personal bias. However, these assessments allow for guesswork to result in high scores. Researchers argue that multiple-choice questions promote memorization and factual recall and do not encourage or test high-level cognitive processes (Airasian, 1994; Scouller, 1998). This is disputed by other researchers who claim how multiple-choice tests are constructed can lead to the evaluation of learning at higher cognitive levels (Cox, 1976; Johnstone & Ambusaidi, 2000). The use of CBM directly helps combat the recognition argument and the guesswork done by students as assessing their confidence for every answer they provide is required (Gardner-Medwin, 1995; Hassmén & Hunt, 1994; Nicol, 2007).

Coupling multiple choice assessments with CBM it is possible to give students relatively basic tests without seeming condescending to high performing students, who get answers mostly correct, while lower-performing students can answer questions initially with low confidence, then increase confidence as they repeat exercises in an area (Gardner-Medwin, 1995). By having the option to assess the basics with CBM, several of the traditional problems with multiple choice and true-false assessment creation vanish. CBM assessments ultimately allow multiple-choice tests to be more accurate and comprehensive in measuring the knowledge of the test taker, while at the same time, allowing test takers to express their doubt and certainty about the answers (Hassmén & Hunt, 1994).
According to research, assessment in education is essential for learning in students. It shapes the learning experience and influences student behaviors. Students take notes, participate in lab activities, and complete homework but often do not fully engage with the content. Only when faced with assessment opportunities do the majority of students seriously engage with the content (Bloxham & Boyd, 2007, p. 3). CBM offers an opportunity to assess one's knowledge during the learning process.

**Confidence-Based Marking Assessment**

In the past, students are familiar with tests, other than CBM assessments, that awarded negative points for incorrect answers. For example, the ACT and SAT both at one time had a reduction in points for incorrect answers. The penalty attempted to reduce guessing on those tests. However, this led to a strategy of students omitting answers if they were not confident about the accuracy. Without negative marking, it is rarely rational to omit an answer. Negative marking for multiple-choice tests has also been proposed as a method for reducing the guessing method, increasing the reliability of the data (Davies, 2002).

When students make a guess, it skews valid exam data with random answers. Students will occasionally end up being correct on a guess without genuinely knowing the correct answer. Guessing does not show subject knowledge. The ability to justify an answer is one of the essential elements in an Aristotelian definition of knowledge, but this skill is often neglecting in students who prefer rote-learning. (Gahan & Gardner-Medwin, 2003; Nicol, 2007). When faced with the likelihood of losing significant points, the necessity of making a subjective confidence judgment may trigger a range of processes including reasoning, monitoring, reflecting and evaluating (Issroff & Gardner-Medwin, 1998). Through continued interaction
with CBM, the metacognitive process in students shifts from unsure and guessing for an answer to the evaluation of one's knowledge.

CBM is not a measure of the students' confidence in themselves, but confidence the provided answer is correct to the assessed question. Students must employ real knowledge, not guesswork, to succeed. This realization that sound knowledge, not ideas, is itself a step toward ensuring academic success and the building of self-confidence in students (Gardner-Medwin, 2006).

**Academic Benefits**

Educators expect students to have the ability to assess the quality of their answers. However, students are often not shown how to do this, nor are they provided any feedback on how to improve on their abilities. For each question, a student is expected to have a mental decision tree to decide if their answer is correct or not. Student confidence falls into several various categories of uncertainty. These levels of uncertainty are end-capped by knowledge and misconception (Davies, 2002). As students answer questions, they often employ some amount of guesswork. A student's decisions, as well as latency, speed, vigor, and smoothness of responses, may be directly related to the self-assessment process (Hunt, 1982). Confident belief in wrong answers is far worse than the recognition one does not know the answer (Gardner-Medwin, 1995).

While taking a confidence-based marking (CBM) assessment, the penalty incurred with a confident wrong answer is often an unwelcome discovery to an unprepared student. The penalty triggers reflection about the reasons for the error and more attention to the explanation (Gardner-Medwin, 2006). Misconception and unreliable knowledge, or lack of
awareness of one's incomplete or incorrect knowledge, can be a sizable obstacle when trying to layer new concepts on inadequately acquired basic comprehension. Tightly or confidently held misconceptions present an obstacle to learning for both student and teacher (Gahan & Gardner-Medwin, 2003; Gardner-Medwin, 1995, 2006). Educators attempt to correct misconceptions without inadvertently strengthening the misconception. The use of CBM can help identify misconceptions. This identification is critical because misinformation is particularly hard to rectify for future growth (Khan et al., 2001). Questions identified as eliciting confident wrong answers can isolate areas where students have significant misconceptions (Issroff & Gardner-Medwin, 1998). Once identified, more targeted intervention could be done to assist the students in the content.

CBM puts students in a position where they risk earning a low score for misplaced confidence. High scores result from the selection of correct answers with high confidence. However, student academic ability will suffer unless students develop the introspective skills which lead to correct confidence judgment or the likelihood of selecting a correct or incorrect answer. Learning these skills can be more valuable than the content in some regards. CBM experiences can help students to develop careful habits of thought and to identify when their knowledge is tentative or subject to critical misconceptions (Gardner-Medwin, 1995).

In the classroom, instruction is intended to have long-term effects on students that last months or years. Given this, achievement testing is often done immediately after instruction, but doing so tests the ephemeral, or short-term, knowledge. Through CBM, the essential skills and knowledge can be tested, allowing educators to determine how well students have learned those essential skills (Ahlgren, 1969). The higher the CBM score, the more confident the
student is in their deep understanding of the content. Becoming a lifelong learner is not solely memorizing copious amounts of information. It is also the acknowledgment of not knowing and the presence of mind to seek out credible sources of information when possible misunderstandings are present. Metacognitive research has shown that metacognition is an important predictor of academic performance; students are able to effectively distinguish information they know and do not know are more likely to review and retain new information (Cubukcu, 2009).

With current educational philosophy moving away from the traditional one-hundred-point scale and letter grade system to standards-based systems, students are now provided the opportunity to look at their knowledge introspectively. Formative assessment, reteaching, and reassessment are the very core of a standards-based framework. Students can learn a subject at their own pace as long as the final result is comprehension. Initially performing poorly on an assessment is ultimately meaningless if students can reassess. CBM can further provide an opportunity present in a standards-based framework. It can help students identify when they have only a tentative grasp of a subject, and it encourages the habit of re-reading the question and reconfirming chosen conclusions (Gardner-Medwin & Curtin, 1996).

When a form of metacognitive training is done, even for a short time, it is shown to improve performance considerably (Cubukcu, 2009). Developing a habit of reviewing in a system where reassessment is the custom would only benefit the students. The ability to indicate confidence or reservation about opinions to others, explicitly or through body language, is an essential skill to possess (Gardner-Medwin, 2006). At the core, this all comes down to more of a question about why confidence marking is not more commonly used at the
secondary level. CBM helps students learn skills not traditionally taught, and confidence-weighted scores have higher reliability and higher predictive validity than more traditional scoring systems (Ahlgren, 1969).

**Concerns against Confidence-Based Marking**

Research investigating the discrepancy between males and females on standardized testing reveals a disconnect exists between course grades and scores on standardized tests. Typically, girls earn higher grades while boys earn higher standardized test scores (Duckworth & Seligman, 2006). Due to its nature, there is the possibility CBM could expand the gap in gender proficiency of students due to personality traits that align more closely with CBM. However Gardner-Medwin (2006) report that students have rejected this idea. The concern is CBM might be disadvantageous to timid or at-risk students, seemingly more common among female students. At University College London, where CBM has been in place for nearly two decades, no significant gender differences have emerged in data from over 3 million answers recorded on-campus computers and in exams (Gahan & Gardner-Medwin, 2003). The use of self-assessment on multiple-choice assessments closed the gender gap between males and females on SAT questions while not lowering the scores of either gender group (Hassmén & Hunt, 1994). Research also shows personal interest plays a factor into self-regulation of knowledge. Students with greater personal interest in a topic and those who view the activity as important or useful are more likely to use adaptive self-regulatory strategies (Cubukcu, 2009). The value of CBM in medical education has been questioned in the past as the tests are often factual recall and do not assess the application of knowledge for problem-solving settings. Medical students often already excel in a learning environment which can alter the validity of the usefulness
outcomes (Khan et al., 2001). However, the data collected along with student feedback shows CBMs continued value.

Another concern is the perception that CBM only benefits students who are already confident in their knowledge. It is essential to recognize the objective of confidence-based marking is not to reward or discourage self-confidence. The aim is to encourage reflection, self-awareness, and the expression of (Davies, 2002; Gahan & Gardner-Medwin, 2003). Making this distinction during CBMs introduction is fundamental.

A further avenue for the argument against CBM is the possibility it will be ineffective at the secondary level. Studies show CBM guides the practice of reflecting upon one's answers and produces useful and practical skills for student growth. With substantial use in medical programs, it is suggested medical students are rarely wrong in their confidence judgments overall (Gardner-Medwin, 1995). At the University College London, CBM coursework is routine, and students have an extensive online practice before use on formal tests (Gardner-Medwin, 2006). Students in the programs may already possess the skills required to be successful using confidence marking. Once again, this prompts the need for more research to be conducted about its use and effectiveness at the secondary level to conclude if it is a viable tool to improve those skills.

**Learning Management Systems**

Technology-enhanced learning in the classroom has become pervasive in education ranging from services such as Google Classroom to Learning Management Systems (LMS). An LMS is a web-based system allowing educators and students to share instructional materials, create class announcements, administer course assessments, submit and return course
assignments, and communicate with each other online (Lonn & Teasley, 2009). Web-based learning tools, like an LMS, can be used as a catalyst for self-reflection and help facilitate change from passive to active learning stimulating students to work on self-assessment. Lecture material can be supported by web-based assessment exercises which provide feedback on the individual level of understanding. Classroom assignments and assessments are now active learning opportunities in the classroom through the use of an LMS (Herse & Lee, 2005).

A common and open-source LMS is Moodle (https://moodle.org). Moodle, a course management software, fits the definition of an LMS. Moodle, an acronym, stands for Modular Object-Oriented Dynamic Learning Environment. This specific LMS, developed under an open-source software project, is powered by the Moodle community (Brandl, 2005).

**Standards Referenced Grading**

In education, there is a growing shift from traditional grading scales to either standards-based or standards-referenced systems. Educational standards of what students are expected to know during a course or in a grade band are now widespread. The primary goal of a standards-based system is for all students to meet the same set of standards. Each student is judged on the same set of criteria used to define proficiency (DuFour et al., 2016, p. 115; O’Connor, 2010, p. 2; Wormeli, 2006, p. 97). A standards-based system clearly defines the difference between formative and diagnostic tools and summative assessments (O’Connor, 2010, p. 107).

In the current K-12 education setting, teachers work in professional learning communities (PLCs) to develop standards alongside formative and summative assessments. The creation of these ongoing diagnostic formative assessments tools is one indicator of a high
functioning PLC (DuFour et al., 2016, p. 136). The formative assessment gives students the opportunity to prove what they have learned as well as give them an opportunity to identify where they need to improve their learning. These formative assessments are where the majority of the learning happens, an opportunity to assess themselves without the repercussions and high stakes of a summative assessment (DuFour et al., 2016, p. 135; O’Connor, 2010, p. 109; Wormeli, 2006, p. 28). Through the use of CBM, students are provided the best of two opportunities. They receive a metacognitive experience as well as an immediate feedback formative assessment of their learning.

**Theoretical Framework**

The framework for this study is a metacognitive theory of learning and behaviorism. Successful metacognitive attributes lead students to become successful lifelong learners. It is directly associated with intelligence in a subject area. Metacognition refers to higher-order thinking, where active control over the cognitive process is present. Students who possess this ability, to any varying degree, can monitor their comprehension and evaluate their success toward any presented task (Livingston, 2003; Schraw et al., 2006). Defined as "thinking about thinking," it is significantly more complicated. In this context, metacognition is both knowledge and regulation (Brown, 1978; Flavell, 1979; Livingston, 2003; Veenman et al., 2006). The implementation of confidence-based marking (CBM) into Physical Science courses tasks students to attempt regulation of their knowledge regularly. Results on CBM assessments yield insight into how well developed a student’s metacognitive regulation is.

Metacognitive knowledge and experiences often overlap and directly affect each other. A student’s metacognitive knowledge base is the reference base when an answer needs to be
evaluated, revised, or abandoned. The larger the knowledge base the student possesses, the more avenues are available when attempting a task. These can be brief or long in length as well as simple or complex in design. The underlying similarity between all tasks considered metacognitive experiences is they occur when the task demands careful or highly conscious thinking about the exercise (Flavell, 1979; Livingston, 2003). All experience is relatively significant as it impacts the knowledge base. More experiences can lead to the creation or abandonment of goals and the addition or deletion from the knowledge base itself (Flavell, 1979).

When students complete a CBM assessment, they are undergoing a metacognitive experience where there is an opportunity to assess their knowledge base. Once the experience is over, addition or deletion from the knowledge base can occur. Research shows an online assessment of metacognitive knowledge yields more reliable results since the assessment of student confidence occurs at the time of the cognitive task. Presenting the assessment after the cognitive task reliability of the results as it is not during the metacognitive task (Veenman et al., 2006).

CBM also taps into research done into behaviorism. Each assessment task, while invoking a metacognitive response from the students is also grounded in the learned behavior from the environment the student is in. Treating the student and the task as the subject and the answers to the assessment as the outcome, we can alter the environment to begin to elicit behaviors that lead to higher achievement (Skinner, 1957). Through the use of CBM, students are placed in a situation loosely based on cause and effect. In this case, misplaced confidence, or lack of ability to evaluate the metacognitive based results in negative scoring. Operating
under the pretense that the effects are direct results of the cause, then negative scoring is a
direct result of misplaced confidence (Skinner, 1985). Understanding this cause and effect
relationship of student behaviors provides information to help educators alter the environment
and begin to alter student behaviors. This use of operant conditioning, both positive and
negative reinforcers, helps students gauge both the breadth and depth of their knowledge base
as well as their ability to judge the quality of the base (Skinner, 1968, 1985).
Chapter 3: Methods

In order to conduct this study, there were many parameters that needed to be determined. These parameters were established in effort to answer the following research questions.

Research Questions

This study will answer the following research question(s):

RQ 1) Do unit exam scores for an electricity unit in a high school physical science class improve after the implementation of confidence-based marking in routine learning management system (LMS) delivered practice assignments and coursework?

RQ 2) Do confidence-based marking scores on a practice test for an electricity unit in a high school physical science class provide educators any predictive abilities to anticipate student achievement on the unit exams?

Participants

The participants for this study were first-year high school students enrolled in physical science at an average-sized (approximately 1,200 students) Midwestern High School (MWHS). Students were presented with three options when enrolling for a science course at registration. The courses offered were biology, physical science, or Earth and space science. Each student self-selected which course they wanted to take. Having a choice of courses leads to a moderately random assortment of students of varying academic abilities in each class period. Some external factors, math scores, fine arts courses, etc., do play a background role in the scheduling for students. Physical science is open to students in 9th through 12th grades. Due to the differences in age, course schedules, and other possible academic factors, data analyzed for
this study consisted of only 9th graders enrolled in physical science in the researcher's class periods. For this study the following information was not collected: student's names, gender, age, race, other identifying information.

At MWHS, freshman students are divided into two separate houses designated as red and blue. With this separation, the students are likely to have a similar set of four teachers for their core content classes. The separation is not completely accurate and there are students who are “cross house” and have teachers from both sides. The houses are separated in a way so collaboratively taught courses for special education are often on one side of the house with EL (English learner) taught sections are on the other. Students with accommodation in the form of IEPs and 504s are often enrolled in Earth and Space science. These scheduling choices also reduce the differences in student academic abilities enrolled in the researcher’s physical science classes. The students are pulled from the “middle road” of students. Remedial math sections are taught on the other side of the house as the researcher, and those students are not often enrolled in the researcher's classes.

Site

At MWHS, the physical science course is one full semester of conceptual and algebra-based physics followed by a full semester of introductory chemistry. This course is presented in the course handbook as a direct preparation course for the offered chemistry courses at MWHS. Students who qualified as participants in the study were those taught in the classroom of the researcher. Data from students in previous academic years of this course served as the control group as the practice assignments, notes, labs, practice tests, unit exams, and final exams were identical to the previous year. The changes from the control to the experimental
group is only the students themselves. These were still a randomly scheduled into the researcher’s classes from the 9th-grade population at the school.

The site itself is classified as a large 4A high school in the central Midwest. In the 2018-2019 school year, there were 1,309 enrolled students. The school is located in a large suburb according to the National Center for Education Statistics. The student enrollment for the 18-19 school year was 74.79% White, 6.57% Black, 8.4% Hispanic, 5.5% Asian, 0.76% Native Hawaiian/Pacific Islander, 4.43% Two or More Races (National Center for Education Statistics, 2019). The students enrolled in the physical science course reflect the ethnic breakdown of the school as a whole but will ethnicity not be examined in this study. The data here is only provided to show insight into the students who were enrolled in the classes and thus participated in the study. However, personal identifiers were not utilized in this study.

At MWHS there is a slow transition to adopt principles of standards referenced grading practices in the classroom of study. The school has adopted policies under the title *Grading for Learning*. With the adoption of standards and the slow transition from traditional grading systems to a scale referenced to standards the change to year-end targets is also present. This comes through at the high school level as the ability to retake and redo practice work as well as assessments. Grades are meant to be indicators of current learning and are used as a communication tool to show how students are progressing on each standard. Students have opportunities to demonstrate their learning for each reporting topic using a variety of classroom and common assessments (Carlile, 2017). Within our school specifically we have received guidance that all students need to be able to complete assignments in a time frame when they learn the material. As part of the PLC process, it has also been decided that students
can retake assessments to show their learning has changed since the initial attempt (personal communication, August 2019). Within these policy changes comes changes to the school culture as well as changes to how students respond to learning. While these are not within the control of the researcher, they are significant enough to be noted. The mindset, work ethic, and responsiveness of the current high school student needs to be framed within the current education landscape.

Data Collection Procedures

Data for this study was collected using the LMS Moodle. Within Moodle, educators can create and upload questions of various types. Scores for practice assignments, consisting of mostly calculate short answer questions, are collected automatically. Within the statistics of each student's scores, both a CBM and a raw score are recorded to be analyzed. Students were required to complete all CBM-related assignments within this physical science course. No extra assignments were created, and students were not judged directly on the CBM scores. This was done as part of the PLC determined process. All assignments across the course, regardless of teacher, must be scored the same. Thus, preventing the CBM score from being used directly. Throughout the semester, students completed the routine practice assignments as part of their learning process. At the beginning of the unit students were introduced to CBM. This included what CBM was and what information can be revealed about their knowledge. Students were required to earn a CBM score of 250% (25/10) to earn credit for an assignment during the unit. The LMS used to administer and collect data regularly presents scores to students in percentages. The data collected and analyzed was all in the form of points, but as students were most familiar with percentages those were used and communicated during the learning
process. In previous units, outside of this study, students needed to earn 90% on practice assignments to receive credit for them. The selection of 250% mirrors this need for a high score but adds the layer of CBM scores as well. Even though students could go back and attempt any practice assignment as many times as necessary or desired they were still encouraged to earn the highest CBM score possible. The ability to redo assignments is in line with PLC and school policies.

Unit exams pulled questions from the same question bank as the practice assignments. The differences from the practice assignments and the exam were two-fold. First, the practice assignments were targeted to a specific essential or proficient skill and contained ten questions. The exams covered a whole learning target, consisting of several basic and proficient skills as well as two to three dozen questions. Unit exams were not CBM-style questions. Within the quiz module itself, practice assignments and practice tests provided immediate feedback CBM style questions. Students answered each question, selected their level of confidence (CBM) and submitted the question to see the score. Unit exams remained unchanged from the previous year. Students were able to check their answers after each question using the adaptive feedback functionality in Moodle. The unit exam was not CBM style questions. This allowed the scores on unit exams from both academic years to be directly comparable.

The final aspect of data collection is a result of the course procedures. Following school and district policies in physical science, students have unlimited attempts on practice assignments. During this study students needed to earn the outlined CBM score to earn credit for the assignments. This is done to keep in line with previous units where required scores were
in place. This was also done provide incentive for students to take the CBM scoring more seriously.

For this study, practice tests were analyzed for their predictive data for unit exam success. Only the first attempt on the practice test was looked at. Following course policies students could make more attempts on the practice test. All students were provided an opportunity, in the classroom setting, to take the practice test. This attempt is where the data collection came from in regard to practice tests. Per school policy, students are guaranteed retakes on classroom unit exams. In this study, only the first attempt on the unit exam was collected and analyzed at for data.

**Researcher Role**

During this study, the researcher served as the students' teacher. With a quantitative study, the researcher should typically attempt to separate oneself from the participants to prevent bias in the study. Ultimately, being a direct part of the participants in the study could alter the data and consequently, the results (Efron & Ravid, 2013, p. 57). Being the educator, being removed from the study, and the participants was impossible. However, to help reduce any possible impacts on the participants, the data were collected with the LMS Moodle, not by the research directly. The earned CBM scores did not directly reflect in student grades but was used as a threshold score as previous detailed for students. The CBM data were not analyzed until after the students completed the semester, and final grades were submitted. By allowing the data collection to happen in the background, the educator's role is still focused on the educational aspects of the course and ultimately, the achievement of the student. When the time came to gather the data from Moodle, the research did not pull and compile the data. The
school has an employee who serves as a technology manager for the site. He pulled the scores and removed all identifying information before sending over the data. Each student’s name and any other information was replaced with a simple whole number scheme. The individual also selected a random assignment to sort all data by, in an effort to change the order of scores around, before they assign the numbers. This was again done to provide more randomization to the data to keep student identification impossible.

**Instructional Method Comparative Research**

This study was a quantitative research study designed to collect CBM (confidence-based marking) data from students with the goal of revealing its usefulness in a high school classroom. CBM is a style of question many students never used before. Students were accustomed to multiple-choice questions; they appear on district-wide standardized tests and in their content classes. Multiple-choice questions ultimately provide a simple and effective method to test student knowledge and to score and analyze data. However, those style of questions lack the assessment of student’s metacognition that occurs while answering the questions. Cook (2005) states the presence or absence of a tool to facilitate metacognition is a suitable example of a comparative study examining the impact of instructional methods. During this research, prior student data were used as the control group to compare to collected CBM data. Unit exam scores are the ultimate comparison as the tests given in both years were identical and did not include CBM themselves. The addition of the control group was a key part of this study. Most comparative instruction method research, when using computer-based learning, suffers from a lack of a control group (Cook, 2005).
**Materials**

The materials required for this research study primarily include the LMS Moodle. The quiz module built into Moodle distributed the confidence-based marking (CBM) questions. The module included many response types within several forms of questions. The question types used in this research were multiple choice and calculated short answer with exact matching. All types of questions (i.e. multiple choice, true/false, short answer, calculated, etc.) are supported with automatic tallying and scoring, based on predetermined rating scales (Brandl, 2005). The settings of the quiz module allowed for the questions to be a CBM. These were presented through the quiz module as an assignment allowing for all questions presented to the students to be CBM. Students each used a Chromebook where they accessed Moodle during the school day.

**Data Analysis**

Data collected in the study were designed to answer the two research questions. The first stated question was answered with a "yes" or "no" conclusion backed by qualitative values. The first research question investigated whether the use of confidence-based marking (CBM) had a direct impact on student unit and final exam scores. This question is answered through a comparative analysis between the CBM unit exams and the control data set. During this time practice assignments, notes, labs, practice tests and unit exams were kept the same as used in the fall 2018 semester. This allowed for the only change to be the introduction of confidence-based marking on the practice assignments. To complete all data analysis collected during this study, SPSS, version 24 was used to run all statistical tests.

To analyze the data from research question one, a Levene’s Test for Equal Variances was first run to determine if there were any significance differences between the control (fall 2018)
and experimental (fall 2019) groups. Assuming equal variances, differences between the two groups were measured using a t-test. The unit exam used in this study contained two separate Learning Targets related to electricity, consequently, the data were analyzed for each learning target. The Levene’s Tests and t-test were calculated for both learning target unit exams.

The effect size calculations difference between independent means is $d$, further referred to as Cohen’s d. For this research the effect size values for Cohen’s d will be small at 0.20, medium at 0.50, and large at 0.80 (Cohen, 1992).

The data analyzed for this study are from assessments related to two different learning targets. Learning Target 1 (LT1) was a conceptual understanding of simple circuits. The basics of LT1 revolved around how current, voltage, and resistance changes in single and parallel circuits based on their construction. Learning Target 2 (LT2) pertains to Ohm’s Law and was written to show a mathematical understanding. Due to the grade level of student and scope of the course this was only the use of Ohm’s Law. For both learning targets, separate practice tests and exams were administered and allowed for individualized comparison to the students enrolled in fall of 2018. Samples of these assessments for both learning targets are found in the Appendix A and B. Samples of both the CBM practice tests and unit assessment for LT1 can be found in Appendix A. LT2 CBM practice tests and unit exam samples can be found in Appendix B.

Research question two investigated whether educators can utilize CBM as a predictor of unit exam achievement. For this analysis, CBM scores from unit exam associated practice tests were broken down into four ranges of scores. As the practice test itself is out of ten points as set by course procedures students earn scores in the following ranges out of a ten-point total. Those are ranges are Negative (-60<=x<0), Low-Positive (0<=x<10), Medium-Positive
(10<=x<20), High-Positive (20<x<=30), where "x" represents the confidence-based score earned by the student. The researcher created the ranges of numbers. These represented several ranges of scores which indicated different levels of success with the CBM questions. The four ranges represent students who through the use of CBM earned a negative score. It was determined these students, regardless of how low the score, represented a range where there are gaps in their understanding of the content. The remaining ranges, (Low-, Medium-, and High-Positive) are all varying levels of success. While the combination of score and CBM certainty can result in a wide range of combinations to receive any score students in the Low-Positive group show they are not as close to a correct and confident understanding as those in the High-Positive group. From there the Medium-Positive is in the middle of these students. Breaking into the ranges from a score standpoint made it easy to categorize students. Students were administered the practice test approximately 2 days before their exam. Due to variations in the school calendar, some students had two days before the exam while other had three. Students were only able to take the practice test with CBM style questioning one time and where able to see the score they earned once they submitted the attempt. After the student attempts were broken down into those ranges, an Analysis of Variance (ANOVA) determined if any differences in the responses from each group were statistically significant. For groups that were significantly different, a Bonferroni post-hoc analysis determined which group combinations displayed significance differences. Both the ANOVA and Bonferroni post-hoc were conducted for both learning targets. To follow up with answering if there was any correlation between CBM practice test and unit exam a Person Correlation was run between the CBM practice tests and the unit exams. This was done for both learning targets.
To show the correlation data graphically scatter plots were used the calculation of effect size as well in the form of a $R^2$ value. Cohen (1992) states that while the definitions of effect size were made subjectively. Since 1977 in his paper Statistic Power Analysis for the Behavioral Sciences the definitions have been fixed. They have also been used generally since that time. When analyzing data from both research questions that same values for effect size were used. The values for small and medium were most significant from Cohen’s work. He describes a medium effect size value as one which represents an effect likely to be visible to the naked eye of a careful observer. This differs from the small effect size is noticeably smaller but not trivial (Cohen, 1992).
Chapter 4: Results and Discussion

Participants

Data collection for this research study took place in fall 2019 using the learning management system (LMS) Moodle. Fall 2019 included the introduction of confidence-based marking into the unit design for the electricity unit. All students were in the 9th grade and enrolled in physical science. More students met the criteria for inclusion in the fall of 2018 (n = 93) than students from the fall of 2019 students (n=66).

Two students in the control group (fall 2018), for learning target two (LT2), had outlier scores that were far from the mean scores. Two students scored 0/10 for the assessment while next lowest value was 5/10. These two scores were left in the control group for the following reasons. First, the students were part of the control group and represent the entire class from which data were collected. Second, being part of the control group, it was determined that removing them from the data would shift the focus from the purpose of this study to determine how the implementation of CBM would alter exam scores. As the experimental group is the focus, comparing their scores to an unmodified control group was determined to be important. Learning target one (LT1) for fall of 2018 had no outliers in the data.

Data Removal in Fall of 2019

Data from seven students in the experimental group (fall 2019) contained one or more missing data points. The decision whether or not to exclude an individual student were made independently for each research question and learning target. Research question one (RQ1) is a comparison of unit exam scores between the control (fall 2018) and experimental groups (fall 2019). Two separate analyses were run as students completed two separate unit exams, one per learning target assessed. Four students did not complete the practice test for LT1 in fall of
Results comparing student scores of those who did not complete the practice CBM test to those who did are displayed in Table 1. There is no collected data to guarantee they experienced CBM during the learning process. The introduction of CBM is what guarantees the students are different from the control group (fall 2018). In addition, both mean scores and standard deviations between these subgroups on assessments (Table 1) show inconsistencies. Students who did not complete the CBM practice test are removed from the remainder of data analysis. Seven students did not complete the LT2 practice test before taking the unit exam. The four students who did not complete LT1 practice test were also all present in the missing data from the seven students for LT2. Again, analysis is displayed in Table 1. These students were removed from analysis for identical reasons.

Table 1

<table>
<thead>
<tr>
<th>Assessment</th>
<th>CBM</th>
<th>N</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning Target 1</td>
<td>Data Absent</td>
<td>4</td>
<td>8.20</td>
<td>.49</td>
</tr>
<tr>
<td></td>
<td>Data Present</td>
<td>62</td>
<td>8.69</td>
<td>1.47</td>
</tr>
<tr>
<td>Learning Target 2</td>
<td>Data Absent</td>
<td>7</td>
<td>9.59</td>
<td>.43</td>
</tr>
<tr>
<td></td>
<td>Data Present</td>
<td>59</td>
<td>9.80</td>
<td>.61</td>
</tr>
</tbody>
</table>

Research question two (RQ2) looked for correlation between the practice test CBM scores students earned and their unit exam score. Similar to RQ1 having data for the CBM practice test is essential for this correlation analysis. Students who presented with missing data were removed from the analysis in RQ2.
Research Question 1

Learning target 1

A total of 155 students completed the unit exam for LT1 (Table 2). A test of variances shows the control and experimental groups were not significantly different ($F_{(153)} = .849, p = .358$). Therefore, equal variances are assumed. Mean score (Table 2) in fall of 2018 for the control group (8.18) was lower than fall of 2019 for the experimental group (8.69). A measured increase (+0.51) was determined to be not significantly different ($t_{(153)} = -1.897, p = .060$).

Table 2
Comparing Unit Exam Scores for Learning Target 1 and 2, 2018 vs 2019

<table>
<thead>
<tr>
<th>Semester</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>$p$</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit Exams Scores for Learning Target 1, 2018 vs 2019</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018 (non-CBM)</td>
<td>93</td>
<td>8.18</td>
<td>1.76</td>
<td>.060</td>
<td>0.24</td>
</tr>
<tr>
<td>2019 (CBM)</td>
<td>62</td>
<td>8.69</td>
<td>1.47</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Semester</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>$p$</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit Exams Scores for Learning Target 2, 2018 vs 2019</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018 (non-CBM)</td>
<td>93</td>
<td>9.53</td>
<td>1.56</td>
<td>.207</td>
<td>0.15</td>
</tr>
<tr>
<td>2019 (CBM)</td>
<td>59</td>
<td>9.80</td>
<td>0.61</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Learning target 2

A total of 152 students completed the unit exam for LT2 (Table 2). A test of variances shows the control and experimental groups were not significantly different ($F_{(150)} = 3.272, p = .072$). Therefore, equal variances are assumed. Mean score (Table 2) in fall of
2018 for the control group (9.53) was lower than fall of 2019 (9.8). The measured increase (+0.27) was determined to be not significantly different \( (t_{152} = -1.266, p = .207) \).

**Research Question 2**

To answer the second research question, data from the practice test and exam data were collected for only the fall 2019 group of students (n=66). For Research Question 2, the CBM practice test scores were compared to unit exam scores to begin to uncover any correlation between them.

**Learning target 1**

For LT1, student CBM scores ranged from negative scores (-60 to 0) up to nearly perfect scores (30). The data were divided up into four ranges (Table 3). LT1 shows average scores increase as the range of CBM practice score also increases. Utilizing and analysis of variance (ANOVA) test (Table 4) it was determined that the changes between groups were statistically significant \( (F_{3,58} = 6.16, p = .001) \). The Bonferroni post-hoc analysis (Table 5) showed only one combination of categories (Negative vs High-Positive) were significantly different \( (p = .001) \).

Given the data from LT 1 are significantly correlated (Table 6), Figure 2 displays the correlation graphically as a scatter plot. A trend line showing the \( R^2 \) value calculated \( (d= .35) \). Utilizing values determined by Cohen (1992) this is a small effect size. This data does a positive correlation between the two even given the wide ranges of scores present.
Table 3

*Correlation between CBM Practice Test Scores and Unit Exam Score – Fall 2019*

<table>
<thead>
<tr>
<th>CBM Range</th>
<th>N</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative (-60 &lt;x&lt; 0)</td>
<td>22</td>
<td>7.86</td>
<td>1.70</td>
</tr>
<tr>
<td>Low-Positive (0 &lt;=x&lt; 10)</td>
<td>17</td>
<td>8.67</td>
<td>1.53</td>
</tr>
<tr>
<td>Medium-Positive (10 &lt;=x&lt; 20)</td>
<td>9</td>
<td>9.20</td>
<td>0.34</td>
</tr>
<tr>
<td>High-Positive (20 &lt;=x&lt;= 30)</td>
<td>14</td>
<td>9.70</td>
<td>0.39</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>62</td>
<td>8.69</td>
<td>1.47</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CBM Range</th>
<th>N</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative (-60 &lt;x&lt; 0)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Low-Positive (0 &lt;=x&lt; 10)</td>
<td>7</td>
<td>8.96</td>
<td>1.49</td>
</tr>
<tr>
<td>Medium-Positive (10 &lt;=x&lt; 20)</td>
<td>5</td>
<td>9.76</td>
<td>0.33</td>
</tr>
<tr>
<td>High-Positive (20 &lt;=x&lt;= 30)</td>
<td>47</td>
<td>9.93</td>
<td>0.23</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>59</td>
<td>9.80</td>
<td>0.61</td>
</tr>
</tbody>
</table>

Table 4

*Analysis of Variance (ANOVA) Comparing CBM Categories with Unit Exam Scores*

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Learning Target 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>3</td>
<td>31.98</td>
<td>10.66</td>
<td>6.16</td>
<td>.001</td>
</tr>
<tr>
<td>Within Groups</td>
<td>58</td>
<td>100.33</td>
<td>1.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>61</td>
<td>132.31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Learning Target 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>2</td>
<td>5.72</td>
<td>2.86</td>
<td>9.89</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Within Groups</td>
<td>56</td>
<td>16.20</td>
<td>.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>58</td>
<td>21.92</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 5

**Bonferroni post-hoc analysis – Multiple Comparisons between CBM score categories**

<table>
<thead>
<tr>
<th>CBM Practice Tests and Unit Exam Score – Learning Target 1</th>
<th>CBM Level</th>
<th>CBM Level</th>
<th>Mean Difference</th>
<th>SE</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative (-60 &lt; x &lt; 0)</td>
<td>Low-Positive</td>
<td>Low-Positive</td>
<td>-.816</td>
<td>.425</td>
<td>.358</td>
</tr>
<tr>
<td></td>
<td>Medium-Positive</td>
<td></td>
<td>-1.346</td>
<td>.520</td>
<td>.074</td>
</tr>
<tr>
<td></td>
<td>High-Positive</td>
<td></td>
<td>-1.846</td>
<td>.450</td>
<td>.001*</td>
</tr>
<tr>
<td>Low-Positive (0 &lt;= x &lt; 10)</td>
<td>Negative</td>
<td></td>
<td>.816</td>
<td>.425</td>
<td>.358</td>
</tr>
<tr>
<td></td>
<td>Medium-Positive</td>
<td></td>
<td>-.529</td>
<td>.542</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>High-Positive</td>
<td></td>
<td>-1.029</td>
<td>.475</td>
<td>.205</td>
</tr>
<tr>
<td>Medium-Positive (10 &lt;= x &lt; 20)</td>
<td>Negative</td>
<td></td>
<td>1.346</td>
<td>.520</td>
<td>.074</td>
</tr>
<tr>
<td></td>
<td>Low-Positive</td>
<td></td>
<td>.529</td>
<td>.542</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>High-Positive</td>
<td></td>
<td>-.500</td>
<td>.562</td>
<td>1.00</td>
</tr>
<tr>
<td>High-Positive (20 &lt;= x &lt;= 30)</td>
<td>Negative</td>
<td></td>
<td>1.846</td>
<td>.450</td>
<td>.001*</td>
</tr>
<tr>
<td></td>
<td>Low-Positive</td>
<td></td>
<td>1.029</td>
<td>.450</td>
<td>.205</td>
</tr>
<tr>
<td></td>
<td>Medium-Positive</td>
<td></td>
<td>.500</td>
<td>.562</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CBM Practice Tests and Unit Exam Score – Learning Target 2</th>
<th>CBM Level</th>
<th>CBM Level</th>
<th>Mean Difference</th>
<th>SE</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low-Positive (0 &lt;= x &lt; 10)</td>
<td>Medium-Positive</td>
<td></td>
<td>-.803</td>
<td>.315</td>
<td>.041</td>
</tr>
<tr>
<td></td>
<td>High-Positive</td>
<td></td>
<td>-.968</td>
<td>.218</td>
<td>&lt;.001*</td>
</tr>
<tr>
<td>Medium-Positive (10 &lt;= x &lt; 20)</td>
<td>Low-Positive</td>
<td></td>
<td>.803</td>
<td>.315</td>
<td>.041</td>
</tr>
<tr>
<td></td>
<td>High-Positive</td>
<td></td>
<td>-.166</td>
<td>.253</td>
<td>1.00</td>
</tr>
<tr>
<td>High-Positive (20 &lt;= x &lt;= 30)</td>
<td>Low-Positive</td>
<td></td>
<td>.968</td>
<td>.218</td>
<td>&lt;.001*</td>
</tr>
<tr>
<td></td>
<td>Medium-Positive</td>
<td></td>
<td>.166</td>
<td>.253</td>
<td>1.00</td>
</tr>
</tbody>
</table>

* = significant at the p = .05 level

**Learning target 2**

Data for LT2 showed no students earned a Negative (-60 to 0) CBM practice test score, making this data different from LT1 (Table 3). With this learning target all students (n=59) averaged 8.9/10 or above on the unit exam. Results from the ANOVA indicated a significant difference between the CBM categories, and their unit exam scores for LT2

\( (F_{(2, 56)} = 6.16, p < .001) \). A Bonferroni post-hoc analysis only found two category combinations
with significant differences (Low-Positive vs Medium-Positive, \( p = .041 \) and Low-Positive vs High-Positive, \( p < .001 \)).

To determine if CBM practice test scores correlated to the corresponding unit exam scores, there was any correlation present a Pearson Correlation was run comparing the CBM practice test to the unit exam score for both LT1 and LT2 (Table 6). LT1 was determined through the Pearson Correlation analysis to be statistically significant (\( r = .592, p < .001 \)). A similar conclusion is shown in LT2 as there is statistically significant results from the correlation (\( r = .516, p < .001 \)).

Data from LT2 are also significantly correlated (Table 6) with Figure 3 displays the positive correlation between the CBM practice tests and the unit exam. The trend line and calculated \( R^2 \) value (\( d = .267 \)) shown a very small effect size utilizing Cohen (1992) values. These trends are present with small Person coefficients (\( r = 5.92, r = 5.16 \)).

Data returned from LT2 showed many instances where students were earning a perfect score (30/10) on the CBM practice test and then earning a perfect score (10/10) on the unit exam for LT2. To gain more insight into this data Figure 4 displays the data in a bubble chart where the plot location is still CBM practice test score vs unit exam scores however, the size of the bubble indicates the number of occurrences of the data point. This was done to display just how many students scored expectational on both the CBM practice test and unit exams.

**Table 6**

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>( p )</th>
<th>( r )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning Target 1</td>
<td>62</td>
<td>&lt;.001*</td>
<td>.592</td>
</tr>
<tr>
<td>Learning Target 2</td>
<td>59</td>
<td>&lt;.001*</td>
<td>.516</td>
</tr>
</tbody>
</table>

* = significant at the \( p = .05 \) level
Figure 2.

*Scatter Plot of Practice Test CBM score vs Exam Score – 2019 Learning Target 1*

![Learning Target 1: Practice Test vs Assessment Score](image)

- $y = 0.0698x + 8.2743$
- $R^2 = 0.35$

Figure 3.

*Scatter Plot of Practice Test CBM score vs Exam Score – 2019 Learning Target 2*

![Learning Target 2: Practice Test vs Assessment Score](image)

- $y = 0.0342x + 8.9909$
- $R^2 = 0.2667$
Discussion

Research Question 1 Unit Exam Scores

This study examined unit test scores for freshman high school students enrolled in a physical science course once an intervention of confidence-based marking was introduced during a unit of study. Much of the research on CBM comes from post-secondary settings as detailed in the literature review. However, a lack of use in K-12 public education is present. Based on the unit exam scores for students enrolled in fall 2019 compared to fall 2018 there is no statistical evidence to show the CBM intervention made any difference to the outcome and success of students as measured by the unit exam. The lack of statistically significant difference (p = .060) in exams scores suggests that including CBM during the fall 2019 semester did not lead to measurable changes in exam scores. These results are contrary to the findings of Issroff
and Gardner-Medwin (2006; 1998) who found that confidence-based assessments were useful to prompt metacognitive processes for students during the learning process. The results from this study suggest that student learning, as measured by unit exam, was not significantly impacted by the use of CBM during the learning process.

The lack of a significant difference between unit exam scores from 2018 to 2019 indicates that the exam scores must be treated as equal. Given the lack of evidence of change from both the t-test and very small effect sizes, the observed increase in unit exam scores between the control and experimental groups cannot be attributed to any interventions done in this study.

During this study students spent only one unit experiencing and working with CBM as well as taking one-unit exam. Cubukcu (2009) detailed that metacognitive training, even for a short time, can show improved performance. However, the students in this study did not improve their learning even though they experienced the metacognitive training provided by CBM. These findings are contrary to what was found by Cubukcu.

Gardner-Medwin and Curtin (1996) stated that CBM can begin to help students build the habit of re-reading the question and reconfirming chosen conclusions. This study only to look at exam scores for students, to determine if academic achievement improved. This would be an excellent area for further study, if students could be more closely looked at to determine how they are approaching questions after the implementation of CBM.

Before analyzing the data, the classroom teacher felt that the scores had not changed compared to the previous year. The intended measurement of RQ1 was to determine if there were measurable changes in student achievement as measured by unit exam scores. It
appeared students performed as well as they did in fall 2018. Being in the classroom and working with my PLC team through the planning and implementation of the unit, there was no change in time spent on the unit. Based on the fall 2018 calendar, we spent approximately the same amount of time on labs, practice assignments, and total unit time. Soderquist (1936) first reported the total time spent on the exam increased greatly as student interacted with the CBM exam. While not direct data was collected, anecdotally it felt like the total time spent during the end of unit exam was comparable to the previous year. The two-part exam was still given during a regular class day as in the previous year. The use of CBM during this study did not increase the amount of time in any measurable way. How this study differs from much of the research is when the CBM was used. In this study it was on practice assignments and practice tests while in other studies, as with Soderquist (1936), it was done on the exam of the material itself.

LT2 also showed a lack of measurable change from CBM implementation. Unit exam scores from fall 2018 compared to fall 2019 showed a lack of statistically significant change. Looking at both average exam score from the two years this lack of change in mean score is also seen. The mean score in fall 2018 was 9.5 out of 10 while in fall 2019 it was 9.8. A change of only 3%. For exam scores this change did not result in letter grade change either. Anecdotally this could be attributed to the learning target itself. Further discussion on this is found in the limitations.

*Research Question 2 Practice Tests and Unit Exam Scores*

The second research question data attempted to show if there was a way for educators to use the CBM data to predict eventual outcome on the exam. This was again broken down
into two learning targets, similar to data from RQ1. When deciding the ranges of CBM scores to assess all negative scores were pulled together in one group and the scores above zero were broken down into three groups (Table 3).

LT1 data shows an increase in exam score as measured by average when the CBM score on the practice tests went up. However, this leads to a conclusion that as students showed they knew more on the CBM practice test they earned a higher unit exam score. There is no direct data to show this would have been any different if the CBM was not present. The correlation between CBM practice test score and unit exam score was determined to be significant ($F_{(3, 58)} = 6.16, p = .001$). Through the Bonferroni post-hoc (Table 5) it was determined that only when students scored in Negative group compared with the High-Positive, two ends of the spectrum, was there any significance. Figure 2 revealed that even through there was a correlation it was widely scattered data. The $R^2$ value, or effect size, returned was also small.

Cubukcu (2009) detailed that metacognition is an important predictor of academic performance. The ability for students to undergo the metacognitive exercise during learning is more likely to allow them to retain new information. The use of CBM allows educators to determine how well the students have learned the information (Ahlgren, 1969). Combining these two, higher CBM score on the practice test would allow the educator to claim students learned those essential skills. It also shows they are likely to perform well on unit exams, especially when comparing high performing CBM students to negatively scoring ones. This allows educators to claim they would have improved academic performance. This study, however, did not collect enough long-term data from students to determine if the score on the
CBM practice test actually can be a predictor for future assessment. A higher practice test score may just lead to a higher test score.

LT2 data followed the same layout of data analysis. With LT2, no students who took the CBM practice test earned a negative score. This means all students showed more confidence and gave more correct answers than compared to LT1. This learning target showed even less of a change between ranges of CBM scores. Again, the correlations between the CBM practice test and unit exam was determined to be significant ($F_{(2, 56)} = 6.16, p < .001$). The Bonferroni post-hoc analysis (Table 5) two combinations of groups were significant (Low-Positive vs Medium-Positive, $p = .041$ and Low-Positive vs High-Positive, $p < .001$). Figure 3 showed that the scores for LT2 were more clustered near the trend line than LT1.

A large number of students were clustered with CBM scores of 30/10 and unit exam scores of 10/10 for LT2. This is show in more detail in Figure 4, the bubble chart. With LT2 showing such clustering of values toward the very top in fall of 2019 and with mean scores starting fairly high as well in 2018 more research into a possible “ceiling effect” for this learning target. This would be need further investigation to see if it played a factor in the results of this study. Figure 4 helps to display just how many students performed expectational on both the CBM practice tests and the unit exam. This visual can be lost in Figure 3 when all the data points end up on top of each other.

**Using CBM in the High School Classroom**

The ultimate question which needs to be addressed is does it make sense to utilize CBM during the whole year at the high school level. Is this really even worth doing? Both research questions were investigating student achievement as measured by unit exams, and correlation
information coming in the form of CBM practice test scores. Based on collected data, it does not appear to be worth the time to use in the classroom as from the single unit of study no statistically significant changes were found. Gardner-Medwin (1995) discussed the importance of developing careful habits of thought, something which CBM is well designed to assist. Having these habits are critical for knowledge acquisition and use as well as identifying misconceptions. However, both of those ideas require long term use of CBM. This study took such a short look into the use of CBM in the high school setting it may not have had time to make real impacts. Davies (2002) and Gahan and Gardner-Medwin (2003) talk about the aim of CBM is to encourage reflection and self-awareness. Data to determine if those skills were obtained would require further research. Those skills may also take time to reveal themselves in assessment scores. In fact, a better time frame may be utilizing CBM for a whole semester worth of learning and then comparing semester exam scores for two groups of students.

Referencing RQ1, there is no statistical test showing the use of CBM made an impact on the unit exam scores. The slight changes can be anecdotally associated with any of the following factors: different group of students, increased teacher efficiency, high quality materials and classroom activities, retakes and reteaching opportunities in the classroom, or other reasons. With the typical LMS having many other tools and question types accessible to teachers there may be other ways to increase achievement in a more measurable way. Because CBM is not a well-known idea there needs to be time taken to explain the process, need, expected outcome, and procedures to student, parent, administrator, and colleagues. As with anything new for students, it also takes time to understand how and why changes like this can affect their metacognitive processes. This might not be able to be seen in only a single unit, or
even a semester. Changes may in fact take years to reveal themselves. Gardner-Medwin (2006) detail the importance of CBM on achievement and metacognitive gains at University College London. Here CBM on exams has been in place nearly two decades. Data from this long of a time frame has the ability to show gains made by students in the area of metacognition. The students in Gardner-Medwin’s study are also very different from those in this study. At University College London the students are high-achieving university students, enrolled in the medical field. This study looked at a sample of local 9th grade students enrolled in a public school. Beyond the difference in level of academics, age is also a major difference.

RQ2 does bring about a similar outlook on the usefulness in the classroom. While there were statistically significant differences found when comparing groups, the data collected does not necessarily determine if there was a need to use CBM. Students who scored well on a standard practice test may also show significant differences than those who just score low on the traditional grade scale. This would be one area I would absolutely include in further study. Because data were collected using CBM on the practice test, it would be excellent to see how a standard practice compared to unit exam looks. Would the trends and similar correlations seen in this study present themselves if CBM wasn’t implemented but students still completed a practice test as they did in this study?

When looking at established research, Gardner-Medwin (1995) stated during informal reactions students appreciated the use of CBM. “They see it as testing something important as related to their knowledge and they like the option of avoiding negative marking.” Continued use of CBM, even in a high school classroom, may eventually bring out positive results in terms of student academic achievement.
**Limitations**

There are several limitations to this research which need to be addressed as part of the conclusions drawn from the data. The K-12 education setting has an ever-changing landscape. The vision from the school district is Teaching All, Reaching All as well as a focus on all students achieving mastery (“Achieving Mastery,” n.d.). The district this study took place in is transitioning to a standards-reference grading system. With this change comes reinforced efforts to allow students to retake assignments, practice tests, and exams. This reflects the goal of a student achieving mastery of the standard having no restriction to attempts and time. There is also a reduction in importance of deadlines and due dates. While the latter does not directly apply to the study it affects student attitude toward learning when the initial opportunity is provided. With retakes there are always multiple opportunities to show knowledge once it is obtained so careful learning the first time is not necessarily important.

Looking at the practice tests these ideas can show why there does not appear to be a more direct relationship. Students took the practice test approximately two days before the exam. With access to the practice assignments and the ability to attempt them as many times as needed to earn a high enough score for grading there is not a way to determine if student who earned a negative score on the practice test did not go home and complete all the practice work. Due to their being no due dates or deadline for work they are were not required to have the smaller, targeted practice assignments done before the practice test was taken. This leaves room in for students to perform poorly on the CBM practice test and then make adjustments in their level of practice before the unit exam. This in itself shows metacognition but is not measurable.
The practice assignments were all subject to the CBM marking. These were unlocked for students to complete throughout the unit, prior to the exam. With the way the learning targets are crafted along with the practice assignments there is a direct correlation from content standard, to practice assignments, to practice test, to learning target exam. Students are likely to have seen all the questions they can be asked on their exam prior to taking it, or at the very minimum versions of the questions. The way the LMS is built and used allows for a bank of questions to be pulled randomly for practice assignments, practice tests and unit exams. This in conjunction with unlimited attempts on practice assignments gives an even greater chance for students to see the questions and practice them multiple times. This was part of the course where the study took place and was out of the control to be changed for this study. Any student provided this opportunity is likely to be more prepared for the questions when they appear on the unit exam, regardless of the use of CBM on the practice assignment.

The second learning target is a mathematical understanding of how electric circuits work. The leads to students only needing to understand and utilize a singular equation to answer the questions on all three tasks (assignment, practice test, unit exam) in the unit for this learning target. Students in the past, even beyond the years of this study, have performed fairly well as whole on this learning target using a variety of teaching methods. Figure 4 clearly shows how many students performed well on this learning target, both CBM practice test and unit exam. With so many students performing well on both the CBM practice test and the unit exam, it is more difficult to draw any conclusions since variations in the CBM practice test data are not present.
**Further Questions and Continued Research**

To improve the quality of this research and the extent of the data collected there are other factors which could be investigated while the CBM is also being implemented. With the policies in place of retakes and multiple attempts on all assignments, the number of attempts per student per assignment could yield insight into how well students do on the unit exam. Relationships between total attempts by students and unit exam scores may be able to indicate why the use of CBM did not account for any changes in achievement. In addition to total number of attempts an LMS often tracks how long students spend per assignment and even question within an assignment. Looking at this data in relationship to how many attempts could lead to insight into how students are processing the questions and how they react to their scores and answers with the use of CBM. As stated above doing similar comparisons to RQ2 but without the use of CBM to see how it ultimately compares once CBM is implemented would revival if CBM makes an impact and to what degree.

One aspect which can only be spoken to antidotally right now is how quickly students finished the assignments, unit exam, and the unit in general. If time permits comparing the time on assignments, practice tests, exams, and how class time is used from a semester without CBM and one with CBM.

If I was going to personally extend this study or have the opportunity to redo it, I would choose a different unit as a starting place. There are some units which have more challenging content for students. This could provide a better opportunity for CBM to do its intended purpose of having students reflect on their understanding. Cubukcu (2009) shared that even a short-term exposure to metacognitive experiences can lead to learning. Callender (2016) on the
contrary states that multiple opportunities to practice and experience metacognitive experiences may be the best approach to long term improvement. These are conflicting ideas which lead to the same conclusion, long term improvement. This study specifically looked at the short term of one unit. Utilizing CBM for an entire semester and doing a comparison of semester final exam scores would be an excellent way to study the long-term benefits of this inclusion of metacognitive practices. Students may begin to employ real knowledge, not guesswork, to succeed over the long term. Garnder-Medwin (2006) cite the realization that understanding knowledge itself is a step toward ensuring academic success. Enduring success that could reveal itself from long term us and data collection of CBM.

Conclusion

During this study the goal was simple, attempt to determine if the use of confidence-based marking in a high school science classroom would have an impact on unit exam scores. To accomplish this, students were exposed to the use of CBM on practice assignments during the learning process in a unit about electricity. When setting out on this study I wanted to see if I could ultimately provide students a way to be more mindful of their answers to questions, provide them additional layers of feedback via an LMS to improve their achievement at the end. The end product was to improve their grades. My professional goal was to figure out if this tool could be used more widely to improve learning beyond my own classroom.

When the study started, I was able to employ my predetermine methods without hindrance. The LMS automatically changed all practice assignments to CBM style assignments. Students were introduced to the concept of CBM with a simple exploratory activity and then we as a teacher and student group took off through our electricity unit. All the same labs and
assignments were completed as the previous year, keeping consistent with the only change being CBM. The whole study ran smoothly, even the practice tests and into the exam. Data were collected for all students and was going to be compared to the previous group. Students were informed the whole time about the study as well as what the possible outcomes could be. The data were collected completely in the background and I did not see any of the data, in a researcher role, until after the final grades for the semester were input.

The results from the study are fully detailed in Chapter 4 of this research. There were several statistical test run to help determine what changes took place in the data between fall 2018 and fall 2019. The quick summary to determine if CBM made an impact on unit exam scores comes down to the t-test results. For both learning targets the results were not statistically different (p = .060, .207). These were followed up with effect size calculations that came out as small (d = 0.24, 0.15) based on research done by Cohen. Based on the research by Jacob Cohen (1992) a medium effect size, or one that would be clearly visible would be 0.50. The calculated effect sizes were very far from the necessary value. Throughout data analysis there were changes in both standard deviation and mean scores in a positive direction.

What I found through the data analysis is a lack of clarity if CBM made a statistical impact. Based on the research by Cohen (1992) the effect size of 0.24 is not small enough to be dismissed but also not large enough to really be noticeable. The t-tests did not show a large enough change to warrant any discussion as to CBM being the catalyst of change. Overall, scores from the year prior to the study increased during the study however CBM cannot be statistically attributed as the sole reason positive changes happened. There are also several limitations to the data as well as the situation in which the study was conducted. This was
detailed in Chapter 5 of this research, but to summarize, philosophies and practices commonly associated with standard referenced grading may change the way students view and attempt curriculum. The ability to redo and retake falls opposite of the purpose of CBM. CBM is meant to help improve one’s understanding of what they really know while retakes allow for misconceptions and false ideas to be held on for extremely long periods of time. Given the situation I would like to see CBM be used again in more setting at the high school level. Perhaps there are other ways it can be incorporated and over longer periods of time where there could be more time to determine if there is any validity in its ability to be used and results in direct improvement in assessment scores.

Ultimately my personal and professional experience showed that CBM, while it did not make statistically measurable changes, is good for student learning. Davies (2002), Gahan and Gardner – Medwin (2003), Cubukcu (2009), and Ahlgren (1969) all claim variations of the same outcome. The metacognition that students are exposed to, undergo, and experience while using CBM in some way does make a positive impact on their learning. Regardless of how short the experience is, or how well they perform, being provided the opportunity to encourage reflection and self-awareness are academic skills all students can strive to improve.
References


https://doi.org/10.1037/0033-2909.112.1.155


Davies, P. (2001). *Computer Aided Assessment MUST be more than multiple-choice tests for it to be academically credible?*
https://repository.lboro.ac.uk/articles/Computer_Aided_Assessment_MUST_be_more_than_multiple-choice_tests_for_it_to_be_academically_credible_/9490151

https://repository.lboro.ac.uk/articles/_There_s_no_Confidence_in_Multiple-Choice_Testing__/9490067


https://doi.org/10.1023/A:1003196224280


Appendix A: Learning Target 1 – CBM Practice Test and Unit Exam Samples

CBM Practice Test – Learning Target 1

Question 1
Not complete
Weight 1.0

A light bulb/pathway is removed from a 2 light bulb parallel circuit making it a 1 light bulb parallel circuit. What will happen to the current?

Select one:
- a. It will increase
- b. It will decrease
- c. It will stay the same

Certainty:  
- Not very
- Fairly
- Extremely

Check

Question 2
Not complete
Weight 1.0

A circuit has 7 light bulbs wired in series. If the voltage going through the battery is 9.0 V, what is the voltage of the first light bulb?

Answer: 

Certainty:  
- Not very
- Fairly
- Extremely

Check

Question 3
Not complete
Weight 1.0

A light bulb is removed from a 3 light bulb series circuit making it a 2 light bulb series circuit. What will happen to the resistance?

Select one:
- a. It will decrease
- b. It will increase
- c. It will stay the same

Certainty:  
- Not very
- Fairly
- Extremely

Check

Question 4
Not complete
Weight 1.0

A circuit has 6 light bulbs wired in parallel. If the voltage going through the battery is 9.3 V, what is the voltage of the first bulb?

Answer: 

Certainty:  
- Not very
- Fairly
- Extremely

Check
Question 5
Not complete
Weight 1.0
A light bulb is added to a 2 light bulb series circuit making it a 3 light bulb series circuit. What will happen to the resistance?
Select one:
- a. It will increase
- b. It will decrease
- c. It will stay the same
Certainty : [Not very] [Fairly] [Extremely]
Check

Question 6
Not complete
Weight 1.0
A light bulb/pathway is added to a 1 light bulb parallel circuit making it a 2 light bulb parallel circuit. What will happen to the current?
Select one:
- a. It will stay the same
- b. It will decrease
- c. It will increase
Certainty : [Not very] [Fairly] [Extremely]
Check

Question 7
Not complete
Weight 1.0
When measuring resistance, the red wire goes in...
Select one:
- a. COM
- b. 10A
- c. V/R
Certainty : [Not very] [Fairly] [Extremely]
Check

Question 8
Not complete
Weight 1.0
When measuring voltage, the probes go...
Select one:
- a. on either side of the object
- b. across the closed switch
- c. across the open switch
- d. on one side of the object
Certainty : [Not very] [Fairly] [Extremely]
Check
Question 9
Not complete
Weight 1.0
A circuit has 5 light bulbs wired in parallel. If the current going through the battery is 4.5 A, what is the current of the first light bulb?

Answer: 

Certainty: : ☐ Not very ☐ Fairly ☐ Extremely

Check

Question 10
Not complete
Weight 1.0
A light bulb/pathway is removed from a 2 light bulb parallel circuit making it a 1 light bulb parallel circuit. What will happen to the resistance?

Select one:
☐ a. It will increase
☐ b. It will stay the same
☐ c. It will decrease

Certainty: : ☐ Not very ☐ Fairly ☐ Extremely

Check

Question 11
Not complete
Weight 1.0
A circuit has 7 light bulbs wired in series. If the current going through the first light bulb is 8.8 A, what is the current of the battery?

Answer: 

Certainty: : ☐ Not very ☐ Fairly ☐ Extremely

Check

Question 12
Not complete
Weight 1.0
When measuring current, the dial is set to...

Select one:
☐ a. 10A (Blue)
☐ b. 20V (Yellow)
☐ c. 200 Ohm (White)

Certainty: : ☐ Not very ☐ Fairly ☐ Extremely

Check

Question 13
Not complete
Weight 1.0
A circuit has 4 light bulbs wired in series. If the current going through the battery is 8.3 A, what is the current of the first light bulb?

Answer: 

Certainty: : ☐ Not very ☐ Fairly ☐ Extremely

Check
Question 14  
Not complete  
Weight 1.0

When measuring resistance, the dial is set to...

Select one:
- a. 200 Ohm (White)
- b. 10A (Blue)
- c. 20V (Yellow)

Certainty:  
Not very  Fairly  Extremely

Check

Question 15  
Not complete  
Weight 1.0

A light bulb/pathway is added to a 2 light bulb parallel circuit making it a 3 light bulb parallel circuit. What will happen to the resistance?

Select one:
- a. It will increase
- b. It will decrease
- c. It will stay the same

Certainty:  
Not very  Fairly  Extremely

Check

Question 16  
Not complete  
Weight 1.0

A circuit has 3 light bulbs wired in series. If the voltage going through the first light bulb is 3.0 V, what is the voltage of the battery?

Answer:

Certainty:  
Not very  Fairly  Extremely

Check

Question 17  
Not complete  
Weight 1.0

A light bulb is removed from a 2 light bulb series circuit making it a 1 light bulb series circuit. What will happen to the resistance?

Select one:
- a. It will increase
- b. It will decrease
- c. It will stay the same

Certainty:  
Not very  Fairly  Extremely

Check

Question 18  
Not complete  
Weight 1.0

A circuit has 9 light bulbs wired in parallel. If the voltage going through the first light bulb is 7.6 V, what is the voltage of the battery?

Answer:

Certainty:  
Not very  Fairly  Extremely

Check
Question 19
Not complete
Weight 1.0

A light bulb is added to a 1 light bulb series circuit making it a 2 light bulb series circuit. What will happen to the resistance?

Select one:
- a. It will decrease
- b. It will increase
- c. It will stay the same

Certainty: [Not very, Fairly, Extremely]

Check

Question 20
Not complete
Weight 1.0

A circuit has 5 light bulbs wired in parallel. If the current going through the first light bulb is 1.6 A, what is the current of the battery?

Answer: [ ]

Certainty: [Not very, Fairly, Extremely]

Check
Unit Exam – Learning Target 1

Question 1
A circuit has 9 light bulbs wired in parallel. If the voltage going through the battery is 8.1 V, what is the voltage of the first bulb?

Answer: 

Check

Question 2
A circuit has 2 light bulbs wired in series. If the current going through the battery is 2.2 A, what is the current of the first light bulb?

Answer: 

Check

Question 3
A circuit has 4 light bulbs wired in series. If the current going through the first light bulb is 9.4 A, what is the current of the battery?

Answer: 

Check

Question 4
As you remove light bulbs in a parallel circuit, the current...

Select one:
- a. Decreases
- b. Increases
- c. Stays the Same

Check

Question 5
A light bulb is removed from a 3 light bulb series circuit making it a 2 light bulb series circuit. What will happen to the current?

Select one:
- a. It will stay the same
- b. It will increase
- c. It will decrease

Check
Question 6
Not complete
Marked out of 10
In a Series circuit with 3 light bulbs, if the first light bulb has a voltage of 4 V, what is the voltage of the Battery?
Answer: 
Check
Question 7
Not complete
Marked out of 10
A circuit has 1 light bulbs wired in parallel. If the current going through the battery is 6.7 A, what is the current of the first light bulb?
Answer: 
Check
Question 8
Not complete
Marked out of 10
As you add light bulbs in a series circuit, the current...
Select one:
○ a. Decreases
○ b. Increases
○ c. Stays the Same
Check
Question 9
Not complete
Marked out of 10
In a Parallel circuit with 3 light bulbs, if the first light bulb has a current of 2 A, what is the current of the second light bulb?
Answer: 
Check
Question 10
Not complete
Marked out of 10
A light bulb/pathway is removed from a 3 light bulb parallel circuit making it a 2 light bulb parallel circuit. What will happen to the current?
Select one:
○ a. It will decrease
○ b. It will increase
○ c. It will stay the same
Check
Question 11
Not complete
Marked out of 10
A circuit has 4 light bulbs wired in series. If the voltage going through the battery is 7.2 V, what is the voltage of the first light bulb?
Answer: 
Check
As you add light bulbs in a parallel circuit, the current...

Select one:
- a. Decreases
- b. Increases
- c. Stays the Same

Check

A circuit has 4 light bulbs wired in parallel. If the current going through the first light bulb is 4.4 A, what is the current of the battery?

Answer:

Check

One at a time you will be sent to the lab area with a sheet of paper to measure an actual circuit.

In the box below, please tell me what your favorite color is.
### Appendix B: Learning Target 2 – CBM Practice Test and Unit Exam Samples

#### CBM Practice Test – Learning Target 2

<table>
<thead>
<tr>
<th>Question</th>
<th>Not complete</th>
<th>Weight</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Question 1</strong></td>
<td>Not complete</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>A light bulb has a resistance of 0.8 Ohms and 6.7 A of current running through it. What is the voltage across the light bulb?</td>
<td>Answer:</td>
<td>Certainty:</td>
<td>Not very</td>
</tr>
<tr>
<td><strong>Question 2</strong></td>
<td>Not complete</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>A light bulb has a resistance of 0.2 Ohms and is connected to a 8.7 V battery. What is the current running through the light bulb?</td>
<td>Answer:</td>
<td>Certainty:</td>
<td>Not very</td>
</tr>
<tr>
<td><strong>Question 3</strong></td>
<td>Not complete</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>A light bulb has a resistance of 17.6 Ohms and 1024.9 A of current running through it. What is the voltage across the light bulb?</td>
<td>Answer:</td>
<td>Certainty:</td>
<td>Not very</td>
</tr>
<tr>
<td><strong>Question 4</strong></td>
<td>Not complete</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>A light bulb has a voltage of 14.4 V and has 1.5 A of current running through it. What is the resistance of the light bulb?</td>
<td>Answer:</td>
<td>Certainty:</td>
<td>Not very</td>
</tr>
<tr>
<td><strong>Question 5</strong></td>
<td>Not complete</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>A light bulb has a voltage of 1018.8 V and has 203.8 A of current running through it. What is the resistance of the light bulb?</td>
<td>Answer:</td>
<td>Certainty:</td>
<td>Not very</td>
</tr>
</tbody>
</table>
A light bulb has a resistance of 5.8 V and has 1.4 A of current running through it. What is the resistance of the light bulb?

Answer:

Certainty: [ ] Not very [ ] Fairly [ ] Extremely

Check

A light bulb has a resistance of 14.5 Ohms and 86.3 A of current running through it. What is the voltage across the light bulb?

Answer:

Certainty: [ ] Not very [ ] Fairly [ ] Extremely

Check

A light bulb has a resistance of 24.6 Ohms and is connected to a 84.3 V battery. What is the current running through the light bulb?

Answer:

Certainty: [ ] Not very [ ] Fairly [ ] Extremely

Check

A light bulb has a resistance of 8.6 Ohms and is connected to a 164 V battery. What is the current running through the light bulb?

Answer:

Certainty: [ ] Not very [ ] Fairly [ ] Extremely

Check

A light bulb has a resistance of 5.8 Ohms and 10.8 A of current running through it. What is the voltage across the light bulb?

Answer:

Certainty: [ ] Not very [ ] Fairly [ ] Extremely

Check
A light bulb has a resistance of 147.2 Ohms and is connected to a 1397.6 V battery. What is the current running through the light bulb?

Answer: 

Certainty: Not very  Fairly  Extremely  
Check

A light bulb has a voltage of 62.0 V and has 4.0 A of current running through it. What is the resistance of the light bulb?

Answer: 

Certainty: Not very  Fairly  Extremely  
Check
Unit Exam – Learning Target 2

Question 1
A lightbulb has a resistance of 18.1 Ohms and 60.3 A of current running through it. What is the voltage across the lightbulb?

Answer: 

Check

Question 2
A lightbulb has a resistance of 17.2 Ohms and is connected to a 88.2 V battery. What is the current running through the light bulb?

Answer: 

Check

Question 3
A lightbulb has a voltage of 4.4 V and has 0.4 A of current running through it. What is the resistance of the light bulb?

Answer: 

Check

Question 4
A lightbulb has a voltage of 1824.8 V and has 220.8 A of current running through it. What is the resistance of the light bulb?

Answer: 

Check

Question 5
A lightbulb has a resistance of 11 Ohms and is connected to a 9.7 V battery. What is the current running through the light bulb?

Answer: 

Check

Question 6
A lightbulb has a resistance of 4.8 Ohms and is connected to a 13.2 V battery. What is the current running through the light bulb?

Answer: 

Check
**Question 7**
A lightbulb has a resistance of 4.8 Ohms and 12.6 A of current running through it. What is the voltage across the lightbulb?

Answer: 

Check

**Question 8**
A lightbulb has a voltage of 14.2 V and has 3.4 A of current running through it. What is the resistance of the light bulb?

Answer: 

Check

**Question 9**
A lightbulb has a resistance of 179.2 Ohms and 1341.0 A of current running through it. What is the voltage across the lightbulb?

Answer: 

Check