Integrating Self-Reflection Measures to Improve the Chemistry Achievement and Self-Predictive Skills of Learners

Barbara Campbell, Juny Coyoy, Lauren Hicks, & Wes Bullock

East Carolina University

June 2018

Abstract

In the experimental study conducted by Nagel and Lindsey (2018), analysis of the data illustrated that, despite repeated opportunities for students to see how well or poor they performed, there remained a significant difference in gains from low-performers to top-performers throughout the course of the semester. In particular, low-performers showed very little to no growth at all, and one speculation of the authors pointed to the lack of reflection on students' performance. In this study, we propose to test this reflection measure as an intervention itself. A high school chemistry course will undergo a similar clicker activity for the length of the semester, with students being randomly assigned to one of three groups: one group that participates in the clicker activity, receives their performance results, and participates in reflection on how to improve their self-predictive skills; one group that participates in the clicker activity and receives their performance results, but does not get time for reflection; and one group that will participate in the clicker activity, but gets neither their performance results nor time for reflection. At the end of the course, we will analyze the results of the data collected for any trends or significant differences among student sub-groups.

Problem and Literature Review

Metacognition, in the aspect of what we are considering here, refers to how well a student can evaluate his or her own understanding of class content (Nagel & Lindsey, 2015). There are a few studies that have been conducted over the past few years in this arena. One in particular reviewed the metacognitive skill of students in a general chemistry course, asking students to predict their performance on an upcoming exam (Hawker, Dysleski, & Rickey, 2016). Results found that the majority of learners overestimated their actual success.

A more recent study by authors Nagel and Lindsey, published in the Journal of College Science Teaching this year, outlined the results of their experimental study which looked at how integrating a metacognitive practice into a chemistry course would affect their learners' self-reflective skills over the period of a semester (2018). Using clickers as a method for students to respond to how well they expected they could answer a question, and then actually trying the question itself, immediate feedback was given afterwards to allow the students to see if they did or did not gauge their skill level well. Analysis of the results discovered that this particular treatment in itself showed large gains for top-performing learners, but no gains at all for low-performers.

Nagel and Lindsey proposed a few ideas as to why this was the case, and one suggestion in particular was to actually build-in time to discuss the feedback that was given to students. Although the aforementioned feedback on how well students gauged their content skills was given after each clicker activity, no reflection on this data or discussion for how to improve was allowed. With a simple tweak to this activity, Nagel and Lindsey believed this could improve the metacognitive abilities of the low-performers. Our follow-up study, therefore, proposes to introduce this discussion of metacognitive feedback to determine if that will improve the performance for all learners.

Specific Research Question

In carrying out this study, we seek to answer the following questions:

1. How will the performance of the group of learners who use clickers to predict their question success, given feedback and reflection time, compare to a group of learners using clickers that only are given feedback, with no reflection, and to a third group of learners who use clickers to predict their skill, but are given no feedback at all?
2. Is there a significant difference in chemistry content achievement among the three groups of learners for a unit of chemistry?

Rationale

The original study detailed the treatment of introducing clickers for students to predict their ability to answer chemistry content questions, and then actually try the content questions themselves. Over the course of the semester, when clickers were used, feedback was immediately given to participants so they could see how well or poor they could predict what they thought they knew. With little gains among low-performers, Nagel and Lindsey questioned if building in time for discussion of metacognitive skills would make a difference. In our study, we set up a similar approach, but introduce this reflection time as the intervention. We use three groups to accomplish this: one group similar to the original who use clickers and receive feedback only, a second group who use clickers, receive feedback, and are given 15 minutes to discuss and reflect on metacognitive strategies, and a third group who use clickers but are not given any feedback. We will look for how the different participant sub-groups perform against one another.

Participants

The participants that will be involved in this proposed study will be a group of high school students in 3 different eleventh grade chemistry classrooms. All of the students who are in the chemistry classes will be expected to participate in the study. Each class will represent one of the three defined groups for this study. Each class will consist of a diverse grouping of students that is representative of the general school population.

Instruments

This study used the following as instruments to gather data: the pre-test, the clicker system, and the post-test. The pre-test and the post-test both ask specific questions regarding knowledge of a particular unit of high school chemistry. The questions themselves are not identical, only similar in skill and standards, in order to prevent influencing the scores. In addition, both tests are the same in length and are to be administered in the same manner.

The clicker system is synced with the questions appearing in the on-screen presentation, and allows students to respond once the questions have appeared. The first question type asks students to predict their success on a given chemistry content question by responding in one of the following ways:

A. I can confidently answer this question correctly.
B. I am unsure if I can answer this question correctly.
C. I am unable to answer this question correctly.

The next immediate question will be that given chemistry question, with a given selection of four possible answers to choose from. Three chemistry content questions, each preceded by three self-predictive questions, will be asked each day of the chemistry unit. Data is collected as soon as a student selects a response with the clicker, and this information is stored within the system.

Treatment

The setting will be the three aforementioned chemistry classrooms, with chemistry as the focus content being taught. The proposed study will contain a group of students who will receive clickers, get feedback, and then the feedback will be discussed for fifteen minutes during chemistry class. The second group will use clickers, get feedback, but will not discuss the results. The last group will receive clickers, but will not receive any feedback. The second and third group will have 15 minutes for individual informal review.

In the original study by authors Nagel and Lindsey the students received clickers, responded to questions with these clickers and then received feedback and had a discussion about the responses, which is similar to the second group in the proposed study. The difference in our proposed study is having the students be in those three different groups and intervening in different ways. Also, we will be tracking the difference among all the groups to see if students improved after self reflection or not.

Instrument Treatment Linkage

Having students reflect on what they predict their content knowledge to be and seeing an actual measurement of that knowledge will have students thinking about their thinking. Using the clicker system, students will predict how well they believe they can correctly answer a given chemistry content question, using the aforementioned responses outlined above. Then, after the self-predictive question is asked, students will proceed to answer the given chemistry content question. Depending on the group of participants in question, the clicker response system may be used to share feedback on the results to these questions.

Design/Procedure

The study will last for approximately 80 instructional days (one high school semester), with equal amounts of time given daily to chemistry content. The content itself will be similar for all three groups. Each group will start using the clickers at the same time. The first group of students will receive clickers, get feedback, and then the feedback will be discussed. The second group will use clickers, get feedback, but will have time for informal review. The last group will receive clickers, will not receive any feedback, but will have time for informal review (in order to equate instructional time).

Data will be continuously gathered through the clickers and observation logs. The data for this study will come from how the students do on the chemistry clicker questions for the chemistry module. We will look at how the group who did not receive any feedback nor have any discussions, did on the questions and compare their results to those of the groups who got either both feedback and discussion or just feedback (See Table 1).

Table 1: Outline of treatment for each of the three groups.

Group	**Treatment**	**Discussion Time**
One	*Pre-test *Clickers *Feedback *Self-reflection *Post-test	15 Minutes
Two	*Pre-test *Clickers *Feedback *Post-test	(15 minutes self-study)
Three	*Pre-test *Clickers *Post-test	(15 minutes self-study)

Table 2: Outline of data obtained from each of the three groups.

Group	**Data Obtained from Participants**
One	*Pre-test *Clicker self-predictive questions *Clicker chemistry questions *Post-test
Two	*Pre-test *Clicker self-predictive questions *Clicker chemistry questions *Post-test
Three	*Pre-test *Clicker self-predictive questions *Clicker chemistry questions *Post-test

Data Analysis

The data for this study will come from how the students do on the chemistry clicker questions (self-predictive questions and chemistry content questions) for the chemistry module and from the pre-post unit test (See Table 2). Analyzing the results of how each group did on the two-part clicker questions will allow us to look for differences among the groups as a whole, using the mean percent correct for each group. In addition, we can break the results down per sub-group to identify any significant trends there. The pre-test and post-test data will allow us to determine if the treatment had any effect on improving the chemistry content knowledge of the students.

Plan of Implementation

The plan of implementation for this study would have multiple steps, consisting of the following:

- ➢ Participant classrooms will be identified and randomly assigned to the treatment conditions.
- ➢ Pre-test on the chemistry unit is given to students and data collected.
- ➢ Students will be instructed on how the treatment procedures will be conducted (including clicker questions, feedback, and reflection, if necessary).
- ➢ Students will be introduced to the clicker system and shown how to correctly use it.
- ➢ Clicker questions will be presented to students on the established basis.
- ➢ Feedback to the clicker questions will be given after each use (for groups One and Two).
- ➢ Reflection measures are introduced upon return of feedback (for group One).
- ➢ Post-test on the chemistry unit is given to students and data collected.
- ➢ Data analysis is conducted to compare results from the pre- and post-tests among each group.

Research Implications

This study would help instructors gain insight into how giving students' feedback and time to self-reflect can help their own learning. It also will provide an understanding of how self reflection affects metacognition.

References

Hawker, M. J., Dysleski, L., & Rickey, D. (2016). Investigating General Chemistry Students' Metacognitive Monitoring of Their Exam Performance by Measuring Postdiction Accuracies over Time. Journal of Chemical Education, 93(5), 832-840.

Nagel, M., & Lindsey, B. (2018). The Use of Classroom Clickers to Support Improved Self-Assessment in Introductory Chemistry. Journal of College Science Teaching, 47(5), 72-79.

Nagel, M., & Lindsey, B. (2015). Do students know what they know? Exploring the accuracy of students' self-assessments. Physical Review Special Topics - Physics Education Research, 11(2).

Possible Publication Outlets

Action Research

Action Research is an international, interdisciplinary, peer-reviewed journal, which is a forum for the development of the theory and practice of action research. The aim of the journal is to offer a viable alternative to dominant 'disinterested' models of social science, one that is relevant to people in the conduct of their lives, their organizations and their communities. The journal publishes quality articles on accounts of action research projects, explorations in the philosophy and methodology of action research, and considerations of the nature of quality in action research practice. "Action Research offers a greatly-needed forum at a time of growing recognition around the world that engagements between researchers and practitioners are central to generating both new knowledge and innovations in practice relevant to many critical problems." L. David Brown, Harvard University, USA

SAGE Publishing. (2016). Action Research. Retrieved from https://us.sagepub.com/en-us/nam/journal/action-research

Educational Researcher

Educational Researcher (ER) publishes scholarly articles that are of general significance to the education research community and that come from a wide range of areas of education research and related disciplines. ER aims to make major programmatic research and new findings of broad importance widely accessible.

SAGE Publishing. (2016). Educational Researcher. Retrieved from https://us.sagepub.com/en-us/nam/journal/educational-researcher#description

American Journal of Community Psychology

American Journal of Community Psychology published in association with the Society for Community Research and Action: The Division of Community Psychology of the American Psychological Association offers quantitative and qualitative research on community psychological interventions at the social neighborhood organizational group and individual levels. Wide-ranging topics include individual and community mental and physical health; educational legal and work environment processes policies and opportunities; social welfare and social justice; studies of social problems; and evaluations of interventions.

ResearchGate. (2018). American Journal of Community Psychology. Retrieved from https://www.researchgate.net/journal/0091-0562_American_Journal_of_Community_Psychology

Appendix 1

With only minimal changes to the course structure, classroom clickers were introduced in introductory chemistry to allow students to regularly compare their perceived abilities with their actual abilities, a measurement also known as calibration. Students used the clickers to provide knowledge judgments, an indication of their confidence in correctly answering questions, prior to actually answering the same assessment item. The clickers allowed responses from this exercise to be viewed by the entire class and permitted additional instruction when the perceived knowledge of the class did not match the actual performance. Overall, introducing this practice resulted in gains in student calibration accuracy from the beginning to the end of the semester. The improvements, however, were not equally distributed across all student groups. Students who performed in the top third of the class appeared to become more confident in their abilities, resulting in greater accuracy in their knowledge judgments. No significant gains were made in knowledge judgment accuracy by students in the bottom third of the class. This study indicates that practice alone does not support improved calibration for students at all ability levels.

Nagel, M., & Lindsey, B. (2018). The Use of Classroom Clickers to Support Improved Self-Assessment in Introductory Chemistry [Abstract]. Journal of College Science Teaching, 47(5), 72-79.

Appendix 2

Metacognitive monitoring of one's own understanding plays a key role in learning. An aspect of metacognitive monitoring can be measured by comparing a student's prediction or postdiction of performance (a judgment made before or after completing the relevant task) with the student's actual performance. In this study, we investigated students' postdiction accuracies for a series of exams within a two-semester general chemistry course. The research questions addressed include (1) How accurate are general chemistry students at postdicting their exam scores? Are there gender differences in postdiction accuracy? (2) How do general chemistry students' postdiction accuracies relate to their exam performance? (3) How do general chemistry students' postdiction accuracies and metacognitive monitoring of their exam performance change over time? Results indicate that most general chemistry students are not accurate in their exam score postdictions and that, consistent with research conducted in other domains, higher-performing students make more accurate postdictions than lower-performing students. In addition, although students who were new to a general chemistry course appeared to improve in their metacognitive monitoring on the second course exam compared with the first, monitoring did not significantly improve after that initial adjustment. Given the importance of metacognitive monitoring for student learning of chemistry, these findings suggest that further research and development of interventions to improve the metacognitive monitoring of introductory chemistry students is warranted.

Hawker, M. J., Dysleski, L., & Rickey, D. (2016). Investigating General Chemistry Students' Metacognitive Monitoring of Their Exam Performance by Measuring Postdiction Accuracies over Time. Journal of Chemical Education, 93(5), 832-840.

Appendix 2 (cont.)

We have conducted an investigation into how well students in introductory science classes (both physics and chemistry) are able to predict which questions they will or will not be able to answer correctly on an upcoming assessment. An examination of the data at the level of students' overall scores reveals results consistent with the Dunning-Kruger effect, in which low-performing students tend to overestimate their abilities, while high-performing students estimate their abilities more accurately. Similar results have been widely reported in the science education literature. Breaking results out by students' responses to individual questions, however, reveals that students of all ability levels have difficulty distinguishing questions which they are able to answer correctly from those that they are not able to answer correctly. These results have implications for the future study and reporting of students' metacognitive abilities.

Nagel, M., & Lindsey, B. (2015). Do students know what they know? Exploring the accuracy of students' self-assessments. Physical Review Special Topics - Physics Education Research, 11(2).

Appendix 3

Overview of the Original Research Question

The original study, titled *The Use of Classroom Clickers to Support Improved Self-Assessment in Introductory Chemistry*, relates the experimental process of authors Megan Nagel and Beth Lindsey as they use clickers to evaluate the metacognition of learners in this chemistry course over the period of a semester. Here, they attempted to discover the answer to questions such as "...does student calibration become more accurate over the course of a semester?" and "Are improvements in calibration uniform for students in all performance groups?" The participants (who chose to volunteer for this study) were in the higher-ed setting and enrolled in the Introductory Chemistry course (Nagel & Lindsey, 2018).

How Original Performance was Measured

In order to answer the questions posed initially, the researchers used classroom clickers to receive student input. Three question pairs were most commonly introduced at the start of a class period. The first question of a pair would ask whether a learner felt he or she would get the stated question correct, and then, after answering, would actually be given multiple choice options to see if he or she could get it correct. Later, the class would review the feedback to see how accurate they were in self-assessing themselves. This data was collected throughout the course of the semester (Nagel & Lindsey, 2018).

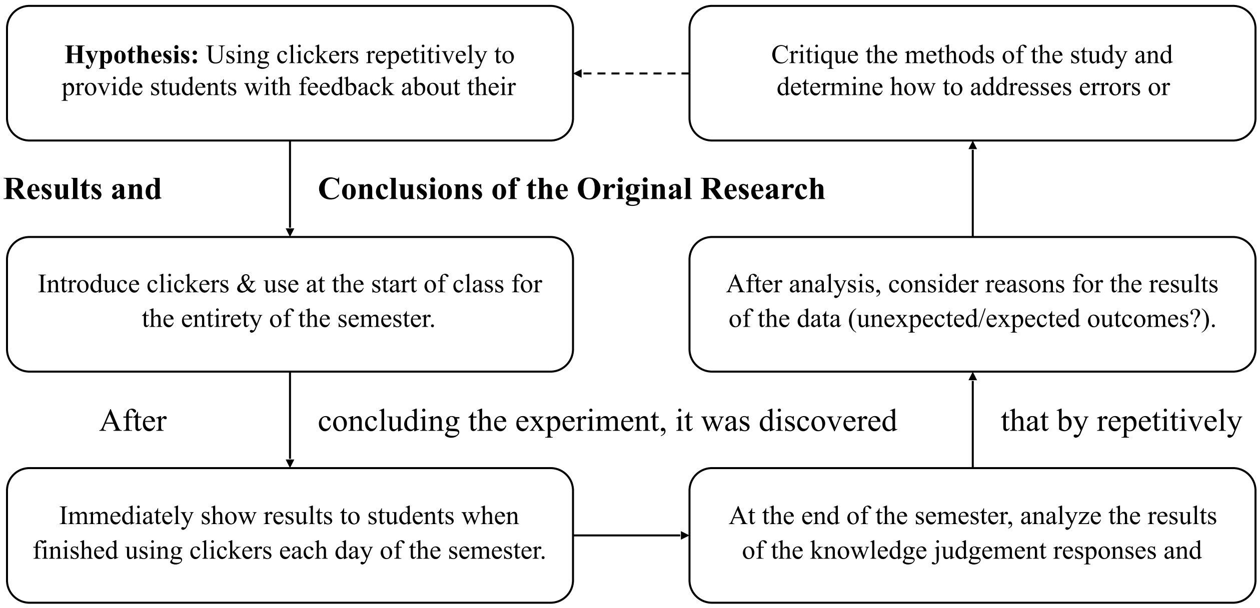
Original Experimental Treatment

The authors proposed that collecting data with clickers at the start of each class period, and then showing the results immediately to learners, would help them reflect on their self-assessment abilities (or lack thereof). Over the course of a semester, the repetition of this practice was expected to assist learners in improving their own "calibration", or how well aligned is a person's ability to judge what he or she knows versus what he or she actually does know (Nagel & Lindsey, 2018).

Design of the Original Research

The format of this particular research followed a design similar to the embedded diagram below. After stating the original intent of the research, the authors introduced their specified intervention (which consisted of using clickers and showing feedback repetitively). Once the data had been collected over the semester, the authors analyzed it to see if any trends had occurred, even amongst certain performance groups (low, mid, & high performers). Once the trends had been realized, the authors speculated on a few reasons for the outcomes and offered suggestions on what a future study expounding on this experiment might entail (Nagel & Lindsey, 2018).

Appendix 3 (cont.)



Results and Conclusions of the Original Research

After concluding the experiment, it was discovered that by repetitively using clickers to show learners their metacognitive abilities, there was some improvement among the class of participants as a whole. However, upon breaking down the data further, it was found that there was a significant improvement in calibration for top-performers, but no improvement for low-performers. In other words, though low-performers consistently observed their lack of self-assessment abilities throughout the semester, this sub-group did not make strides toward increasing their abilities, based on the data collected (Nagel & Lindsey, 2018).

Potential Improvements to the Original Research

As suggested by the authors, using more challenging questions may have resulted in a different outcome for the top performers. In other words, the feeling is that this exercise may have simply helped bring the confidence level up to the performance level of top-performers (with more difficult questions, top-performers would not have simply realized they would get all the questions correct). In addition, I question how the data may have looked different in another course structure. Since repetition was a key to this exercise, how would the data have turned out if the class met more frequently? The authors stated this class met twice a week for 50 minutes (plus a lab section). How would meeting daily have affected results, as repetition would have been far increased? Finally, as suggested by the researchers here and proposed by ourselves, how would metacognitive skills be changed if an intervention was used to provide serious reflection on calibration skills throughout the course of the study? Immediate feedback was provided in the original study, but no time was given to assist learners in considering why they were or were not judging themselves correctly.