Efficacy of Various Formative Assessments on Student Achievement in an Undergraduate Introductory Biology Class

Elena McIntosh

George Mason University

EDIT 590

Abstract

An investigation into the topic of formative assessment reveals gaps in the current research. This research proposal takes into account previous shortcomings of studies on formative assessment and heeds the advice of the writers of a meta-analysis of formative assessment research to shift the conversation from a discussion of *whether* formative assessment is effective to *which types* of formative assessment are the most effective or *what factors* influence the effectiveness of formative assessment (Kingston & Nash, 2011, p. 35). A call for high quality empirical research has been made for the topic of formative assessment. This research will be an experimental study with a randomized to groups pretest-posttest design with multiple variables. The dependent variable is variance in student achievement between a pretest and summative assessments and a final posttest on Biology 101 content. The independent variables are four different types of formative assessments administered to the experimental group only. Results on the pretest and posttest will be analyzed by groups of objectives which align to each of the formative assessments and each of the summative assessments.

Key Terms: Formative Assessment, Assessment and Achievement, Assessment for Learning, Learning and Testing

Research Proposal

Introduction

While formative assessment is generally believed to have a positive impact on student achievement, studies that have been used to draw this conclusion upon closer inspection have been shown to be methodologically weak. The design of the current research varies significantly, making it difficult to replicate or compare results. Much of the research involves instructors observing their own classes, making objectivity difficult. The question of whether formative assessment works is no longer the best question to be asking. Rather, researchers should begin to study the factors influencing the effectiveness of formative assessment. A new approach to research on formative assessment is needed.

In my experience as a teacher, I have felt that timely feedback was essential to the learning process for students. Rather than structure the classroom to have a content delivery focus, I preferred to take on a role of facilitator of learning and focus more on giving students the opportunity to engage with materials that would lead to the desired learning outcomes I have set. Because I take the constructivist and social cognitive perspective that knowledge is created and shared by groups of learners, I believe it is the educator's role to provide the resources and the guidance to structure the knowledge creation of the learners and to influence the learners to engage with the materials. The information that students need to learn must be incorporated into their schemata. This process of transforming external stimuli (information input) into thought and encoding it into long-term memory (i.e., the learning process) is complex and should be studied along with the study of the tools educators are using to influence this learning process.

Thus, the question to ask is: why does formative assessment have an impact on student achievement? Is it because it provides repeated exposure to the content? Is it because teachers design formative assessments that tend to mirror questions on summative assessments (modeling teacher expectations)? Is the level of questioning (knowledge versus critical thinking) a factor in how much a formative assessment impacts student achievement? Does the learning modality (oral versus written) of formative assessment impact student learning? This study will focus on the last two questions in an attempt to determine which factors of formative assessment have the greatest effect on student achievement.

Literature Review

Definitions for formative assessment research. Formative assessment is assessment *for learning*, or in support of learning. It is contrasted in the research by summative assessment, which is assessment *of learning*, or of student achievement. By definition, formative assessment should not impact the grade of the student, but should inform the student and the teacher of the student's progress toward learning goals. This definition of assessment from Maria Weurlander, Magnus Söderberg, Max Scheja, Håkan Hult and Annika Wernerson is a synthesis of the definitions of their predecessors in research on assessment:

> Assessment is about making judgements on the quality of students' performance (Knight 2006). It can be used both to summarise students' achievements in order to award some kind of certification (summative assessment) and to give feedback to students in order to support learning (formative assessment). (Weurlander, p. 748)

Formative assessment defined as "feedback" is commonplace in the literature on assessment. The function of formative assessment is to provide the student and the teacher with feedback on the current status of student learning in order to guide the teacher to make instructional decisions

and guide the students to make study decisions. The implications of the feedback can be wide-ranging, from providing insight into the quality of the teacher's instruction, to indicating the effectiveness of the assessment as a measure, to identifying the actual level of learning that an individual student is achieving. According to Keefe and Eplion, "[formative assessment] is used before or during instruction to assist learning and provide specific information about students' strengths and weaknesses to help them and their instructors with the task of teaching/learning" (2012, p. 59). Their study attempted to refute the idea that formative assessment serves primarily as feedback, proposing instead that the primary function of formative assessment is as a tool for intrinsic and motivational change (Keefe & Eplion, 2012). The results of this study did not conclusively defend the hypothesis that formative assessment primarily serves to motivate students to improve their study habits. Thus, the definition of formative assessment as "feedback" will stand for this proposal.

Findings of current research. Some of the studies reviewed for this proposal made claims about the effect of formative assessment on student learning, even when the effect size between the control and study group was not statistically significant. According to Horine, "statistical significance is not needed to have important practical significance. In the end, only the reader can determine what is practical and meaningful in using the results. In this sense, [the reader's] conclusions are more important than those stated by the researchers" (McMillan, 2011, p. 255). The literature reviewed for this proposal shows that researchers have overwhelmingly concluded that formative assessment has a beneficial impact on student learning and motivation to study. Students in several studies reported being more aware of and invested in their learning as a result of the feedback they received through formative assessment (Keefe & Eplion, 2012; Weurlander, Soederberg, Scheja, Hult, & Wernerson, 2012; Winstone & Millward, 2012). Two

studies in particular focused on the motivational and metacognitive impacts of formative assessment, the Keefe & Eplion (2012) study of business school students taking online formative assessments before in-person lectures and the Hudesman et al. (2013) article detailing the results of a program called EFAP-SRL on the scores of associate's degree students in developmental math classes. Both studies claimed that formative assessment had an impact on the students by motivating them to work harder to achieve higher scores.

Methodological concerns of current research. One major area of disagreement amongst researchers of formative assessment has to do with the mean effect size of formative assessment calculated using as many research studies as qualified under specified standards for research. The mean effect size for studies of formative assessment was originally reported to be between .40 and .70 in a study by Black and William in 1998. However, a more recent meta-analysis found a weighted mean effect size of .25 over 13 studies selected for review (Kingston & Nash, 2011). In their meta-analysis, Kingston and Nash reviewed 300 studies and found only 13 studies that contained enough valid data to calculate effect sizes (2011, p. 28). In 2012, Briggs, Ruiz-Primo, Furtak, Shepard, & Yin raised concerns about the methodology of the meta-analysis conducted by Kingston and Nash and attempted to call into question the new median effect size of .25 (p. 13). They criticize the approach taken to select studies, the data analysis conducted on the study results and the experimental design of several of the studies reviewed. The main takeaway from Briggs et al. is that the research studies currently available contain many flaws which could compromise the validity of the effect size number. Factors which could influence student achievement cannot easily be controlled for. The correlation between the independent and dependent variables is difficult to clearly delineate, with factors such as teacher performance, student aptitude and interest, interpersonal relationships between teacher and

student, peer influence, quality of instructional materials/activities, external environmental factors and more, all possibly having an influence on the results. Another very valuable insight from the literature reviewed is the idea that both students and teachers have a predisposition to favor a certain type of assessment. Weurlander et al. call this an "orientation" to assessment. Citing an article by Samuelowicz and Bain from 2002, they state that "teachers' orientation to assessment range along a continuum, from a focus on knowledge retention to an emphasis on knowledge construction and transformation, and is related to their orientation to teaching and learning" (Weurlander, p.748). This orientation can influence student achievement data. In several of the studies, instructors design the assessments, which could allow them to select material covered in the formative assessments for the summative assessment in order to influence the results. "Given the wide use and potential efficacy of good formative assessment practices, the paucity of the current research base is problematic. A call for more high-quality studies is issued" (Kingston & Nash, p. 28).

Lessons learned from current research. The critique by Briggs et al. of the meta-analysis by Kingston and Nash reveals several important factors to consider when designing a study of formative assessment. First, I realized that I should include a pre and post-test in order to compare the student to himself before and after the treatment of formative assessments. Second, I discovered that having a control group with a similar demographic to the experimental group is valuable when drawing conclusions about the effect of formative assessment. For this reason, I have designed the study to include a control group. Third, it is important to define the assessments (outcome measures) clearly and distinguish between the questions used for formative and summative assessments so that the results are not influenced by exposure to similar questions. The assessments designed for this research proposal will be peer-reviewed and

checked to avoid repetition with the summative assessment. Also, the instructor should not be the person to design the formative assessments. Rather, external, objective researchers should participate in the study design and implementation. Finally, and most importantly, I learned that it is time to move beyond the question of whether formative assessment is effective and begin to study which types of formative assessment are most effective. The literature review was most helpful in that it provided me with a new focus for my research.

Research Problem, Purpose and Hypothesis

Problem and Purpose. Which assessment tools are most effective for promoting student learning? This research will attempt to distinguish the effect of different types of formative assessments on learning outcomes for students in a required freshman Biology 101 class. The purpose of this study is to determine which types of formative assessment have the most powerful impact on learning as demonstrated through student achievement. A knowledge level, multiple-choice, written assessment, a higher-order thinking short answer written assessment, a peer conducted oral (knowledge level) assessment, as well as an instructor directed discussion with a peer (higher-order) will be compared to determine which has the greatest impact on summative assessment results for students.

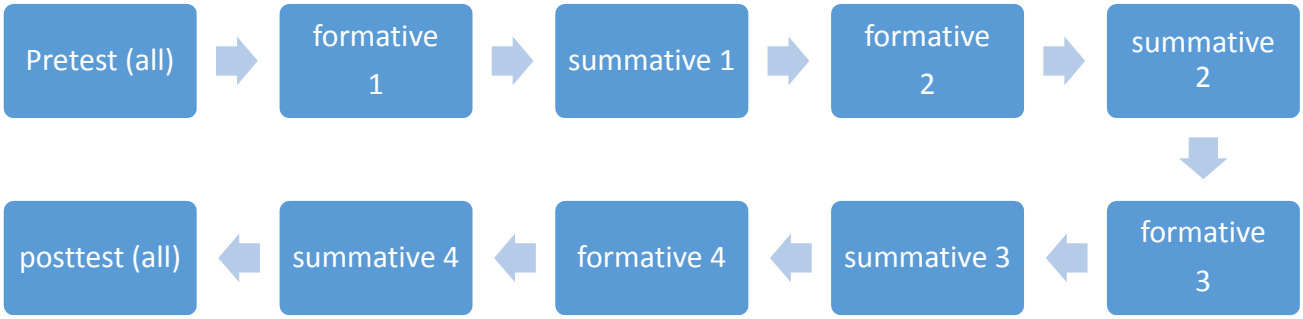
Hypothesis. I hypothesize that students in the experimental group receiving formative assessments will score higher on each of the four summative assessments and the posttest. I believe the results will prove that the effect size of formative assessment is stronger for the higher order thinking assessments than for the knowledge level formative assessments. I also hypothesize that between the two knowledge level assessments, the peer conducted assessment will have a greater impact than the written assessment. Between the two high-order thinking

assessments, the instructor led peer discussion will have greater impact than the short response quiz. The study will confirm that the most effective type of formative assessment is one in which students must produce a verbal response using higher-order thinking skills.

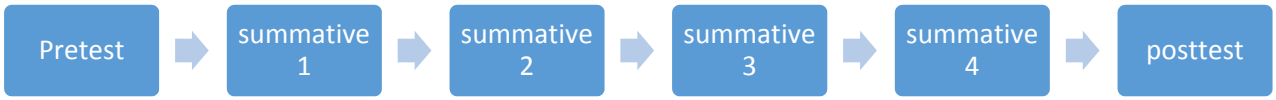
Methodology

Design. This research will be conducted as an experimental study with a randomized to groups pretest-posttest design. However, the study will involve multiple (four) independent variables. In order to maintain the pretest-posttest design, each summative assessment will serve as a posttest for some of the objectives covered in the pretest. The pretest will be comprehensive and will cover all of the objectives for the course. The four summative assessments will each cover a group of content objectives. Group 1 content objectives will be covered on formative assessment 1 immediately preceding summative assessment 1 on that content and so on. A cumulative posttest will be administered at the end of the course for a second source of achievement data. Both the pretest and the posttest will have questions grouped by content that exactly match the groups of content for each of the four summative assessments. Finally, in order to eliminate the possibility of a difference between the content groups in difficulty, a previous year's assessment results on all four summative assessments and the posttest will be compared. Each test will be compared to determine if there is a pre-existing difference in scores between the summative assessments. This pre-existing difference will be taken into account when calculating the final effect size.

Experimental Group Assessment Chart



Control Group Assessment Chart



Participants. Adult college students in a freshman level introductory biology course will be the focus. The students will be randomly assigned one of four sections of this class. Students who have already taken the course or who are declared science majors, but have chosen to start with an introductory class will not be included in the study. Students may elect to take the course in a different semester if the time randomly scheduled does not fit with their schedule. We will rely on student advisors to help explain the option to the students without revealing that the class will be participating in a study.The anonymity of the participants will be protected through the use of a currently established practice for the professor of requiring students to write only their student ID # at the top of the graded assessments. Scores will be recorded by student ID number and then

a code will be attached to each student ID number, to allow data to be stored upon conclusion of the study by the code number only. This will provide long-term anonymity.

The study will compare the assessment results of an experimental group of students taking formative assessments one class period before scheduled summative assessments and a control group taking only the four summative assessments, which will be final and count as their only grades.

Variables	**Attribute**
Student achievement on four summative assessments (dependent)	% scored out of 100
Type of formative assessments administered during the course (independent)	Knowledge level multiple choice (written), higher-order short answer (written), knowledge level quiz conducted by peer (oral), higher-order instructor directed discussion with a peer (oral)

Recruitment and Sampling. Because this study will be conducted with an experimental design, participants in the study will not be aware that they are participating in a comparative study. The total number of students taking Biology 101 during this semester will number

approximately 500, with approximately 125 students in each section. The professor will be aware that he is administering different types of assessment to two sections of Biology 101 than he is to the two other sections of Biology 101.

The target population is college students in their first year of college. The population being sampled from will consist of college students attending a reputable university and enrolled in a non-science degree program, thus taking a required introductory science course. College students are assumed to have high school level understanding of the four core content areas. This Biology course may include content they have already learned, are re-learning or have never learned. Students may have declared majors or may still be undecided. As a result, the Biology skill levels of the students should be diverse. Some may have more knowledge and interest in science than others. The implications of the research should be applicable to adult learners and may also be relevant to high school students.

Setting. This research will be conducted at the University of Virginia within the already existing structures of the University, Biology and Sciences department and the already extant Biology 101 course. The control group and the experimental group will both include one early class, 8:00am start, and one later class, 2:00pm start, to ensure that any differences in achievement that may occur between early risers and later risers will not affect the results of the study. Students will participant in synchronous live lectures with engaging activities and materials as prepared by the Biology professor with no changes to instructional delivery method and style. The formative assessments will take place at the end of class on the day before the summative assessment on the same topic. Students will be given a maximum of 15 minutes to participate in the formative assessments. The professor will conduct the course in every other way just like the previous year's course.

Materials. Formative assessments must be designed and reviewed to be aligned to the course content and the professor's lectures and summative assessments. The researchers will need access to all course content and texts in advance in order to design the formative assessments. The formative assessments will need to be administered in a timely fashion to ensure they are instructionally appropriate for the content covered that day and on the summative assessment the next class period.

Implementation procedures. After reviewing the course content and text, the researcher will design the four formative assessments and have the assessments peer reviewed and after further editing, field-tested by a group of graduate research students. The researcher will train the instructor on the implementation of the formative assessments. Once the instructor is trained on how to implement the formative assessments, the instructor will administer the formative assessments. The control group will be directed to take those same 15 minutes to study their notes from class that day and class will end at that time, allowing both groups to have the same amount of instructional and review time. No other changes will be made to the delivery of instruction. The summative assessments will then be administered on a schedule and the results of each assessment collected by student ID.

Considered threats to validity and reliability. Several presumed threats to reliability have been taken into account, including potential differences in student backgrounds and current levels of achievement, meaning that one group may already tend to achieve higher results on assessments than another group. The potential lecture time differences due to inclement weather or differences in student questioning and participation during lectures is another threat to reliability. Other factors that could influence the reliability of the research include: student attendance, internal factors influencing their motivation, relationship with the professor, study

habits, background knowledge on new material, familiarity with and ease of use of the assessment tool. Variability in the professor's delivery and mood based on the time of day, energy level or sequence in which the class occurs (first class versus last class receiving the same instruction) could also influence the student achievement. These factors will need to be accounted for in the statistical analysis of the data.

Threats to validity could include the differences in difficulty of summative assessments by content group which could naturally tend to yield different student achievement scores. To attempt to address this, the previous year's student achievement data will be requested to discern whether there is a difference in the mean scores for each of the summative assessments before the experimental year. The validity of the conclusions could be threatened if the population is not viewed to be one that can be used to generalize to other populations. Perhaps high-achieving students such as those at UVA would be more likely to be aided by formative assessment than students who are less achievement oriented.

Proposed data analysis. In order to determine the true effect size of the formative assessments on student achievement, various procedures will be implemented to ensure that the data is analyzed accurately and the study conclusions are valid. First, a null-hypothesis shall be established of .05. If the mean summative assessment results for the experimental group are less than 5% higher than the control group, the study will conclude that there is no significant effect. Mean pre- and posttest data for the two experimental and the two control sections will be compared to determine whether there is a pre-existing achievement difference. The mean scores for the four summative assessments will be compared to one another and to the previous year's assessment results to determine whether there is a significant effect in the experimental group that cannot be explained by a difference in the difficulty of the summative assessments. The

mean scores of the pretest and posttest will be compared by objective (content group) with the summative assessments on those objectives to determine whether there is a greater improvement in the experimental group than in the control group for each of the formative assessments administered.

Conclusion

Through this proposed research, I hope to answer the call to provide quality empirical research on the topic of formative assessment. I hope to illuminate the question of which factors determine the effectiveness of formative assessment. In the end, the research should show that the quality of formative assessment matters and that a teacher cannot expect to achieve a huge impact by simply instituting the practice of formative assessment without first considering what types of formative assessment are the most useful. Teaching is an art, not a science, but an investigation of the practices of teachers and the impact of those practices on student achievement can help us become more effective within our own creative approaches.

References

Briggs, D., Ruiz-Primo, M., Furtak, E., Shepard, L., & Yin, Y. (2012). Briggs, D. C., Ruiz-Primo, M., Furtak, E., Meta-analytic methodology and inferences about the efficacy of formative assessment. *Educational Measurement: Issues & Practice, 31*(4), 13-17. doi:10.1111/j.1745-3992.2012.00251.x

Bulunuz, N., Bulunuz, M., & Peker, H. (2014). Effects of formative assessment probes integrated in extracurricular hands-on science: middle school students' understanding. *Journal Of Baltic Science Education, 13*(2), pp. 243-258.

Hudesman, J., Flugman, B., Crosby, S., Issac, S., Everson, H., & Clay, D. B. (2013). Using formative assessment and metacognition to improve student achievement. *Journal Of Developmental Education, 37*(1), pp. 2-13.

Keefe, T., & Eplion, D. (2012). Motivational impacts of formative assessment: Evidence from a college classroom. *Global Education Journal, 2012*(2), pp. 58-91.

Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues & Practice, 30*(4), pp. 28-37. doi:10.1111/j.1745-3992.2011.00220.x

Krumsvik, R., & Ludvigsen, K. (2013). Theoretical and methodological issues of formative e-assessment in plenary lectures. *International Journal Of Pedagogies & Learning, 8*(2), pp. 78-92.

McMillan, J. (2011). *Educational research: Fundamentals for the consumer* (6th ed.). Boston: AB Longman.

Perera-Diltz, D. M., & Moe, J. L. (2014). Formative and summative assessment in online education. *Journal Of Research In Innovative Teaching, 7*(1), pp. 130-142.

Vonderwell, S., & Boboc, M. (2013). Promoting formative assessment in online teaching and learning. *Techtrends: Linking Research And Practice To Improve Learning, 57*(4), pp. 22-27.

Wenjie, Q., & Chunling, Z. (2013). The analysis of summative assessment and formative assessment and their roles in college English assessment system. *Journal Of Language Teaching & Research, 4*(2), pp. 335-339. doi:10.4304/jltr.4.2.335-339

Weurlander, M., Soederberg, M., Scheja, M., Hult, H., & Wernerson, A. (2012). Exploring formative assessment as a tool for learning: students' experiences of different methods of formative assessment. *Assessment & Evaluation In Higher Education, 37*(6), pp. 747-760. doi:10.1080/02602938.2011.572153

Winstone, N., & Millward, L. (2012). Reframing perceptions of the lecture from challenges to opportunities: Embedding active learning and formative assessment into the teaching of large classes. *Psychology Teaching Review, 18*(2), pp. 31-41.