# Practicum Research Proposal:

# Assessing the Effectiveness of Formative Evaluation in Teaching Mathematics Online

Tim Xeriland

Michigan State University

## Abstract

I propose an experiment to test how *formative evaluation* affects performance, interest, and transfer of learning in students studying mathematics in an online collegiate course environment. Formative evaluation is specifically defined in this study as, the evaluation of assessment-based evidence for the purposes of providing feedback to students regarding the rationale of their responses in an attempt to correct *buggy algorithms*, *lack of structural knowledge*, *errors in logic*, and *misconceptions* in their thought processes. Because there is so much ambiguity in the literature regarding the formative assessment concept, the included definition is intended to clarify for the reader the use of the term and that it is not necessary to match others' meanings. Although formative evaluation has been proven to be effective in traditional courses, technological constraints make it challenging to use this approach for online courses. The proposed research seeks to determine if such technological barriers are worth overcoming and have the implication of changing the way certain courses are currently assessed online. Specifically this study will look at a sample of approximately 80-150 community college students working in an online environment where content will be covered in two different modules. One module will be on applying congruent triangles and a second on circles and angle measurements. Each module will be followed by a quiz and the experimental group will receive formative evaluation based on the results. At the conclusion of the treatments a posttest will determine content mastery of the students. In addition, visual matrices problems, not covered in the content, will be included on the posttest to assess transfer of learning. Also, a self-assessment questionnaire will be administered to establish interest gained in the subject matter.


*Key Words:* assessment, evaluation, formative, summative, misconceptions, mathematics

Assessing the Effectiveness of Formative Evaluation

in Teaching Mathematics Online

## Problem Statement

The longstanding belief has been that as long as there is an expert who explains a concept clearly, and as long as there is a student paying attention carefully, that the student will begin to build up the understanding of the expert. Results from educational research tell us that a student may well be paying attention carefully to what is being said, but they are construing it in ways unintended by the expert (Baturo & Nason, 1996). That is, they develop misconceptions, which are when students misunderstand the basic concept of what is taking place (Li & Li, 2008). An important part in remedying this is for teachers to ascertain how the information is being assimilated.

One significant way that teachers can determine how the content is being understood is through assessment and evaluation. The traditional approach of evaluation, however, involves students receiving grades for tests, assignments, and papers, and afterwards instruction moves to a new topic. This is called the summative approach and has advantages because it is relatively easy to create, administer, and evaluate these kinds of tests. The problem is these assessments and evaluations typically fail to offer guidance to students on how their work can be improved. Even if the teacher spends time giving feedback in the form of written comments, as long as a grade is also attached to the evaluation, the comments have been shown to be of no use in helping students improve (Butler, 1988). Dylan Wiliam sums these finding up by saying, "if you are going to grade or mark a piece of work, you are wasting your time writing careful diagnostic comments" (Wiliam, 1999, p. 8). One possible reason for this is because once students receive a final grade, their learned response is to move on to a new topic ignoring all qualitative feedback, hindering their ability to improve on errors. Another possibility is that emotionally the only significant event is the grade and students are not able to focus on the qualitative feedback.

In contrast, formative assessment, which can be thought of as assessment that spurs further learning, (Chappuis & Stiggins, 2002; Waterman, 2010) requires teachers to access the thought process behind students' answers. Teachers can use this information to give students more opportunity to interact with their misunderstandings. Black et al. (2003) reviewed the formative assessment literature and were able to "identify 20 studies that showed that innovations which included strengthening the practice of formative assessment produce significant and often substantial learning gains" (p. 41). Lipnevich and Smith (2009) found that feedback given to students as a result of formative assessment had a positive effect on learning, even when the feedback did not come from the instructor, but was computer generated. Although this approach has been demonstrated to be effective in the classroom it is often superseded by traditional evaluation methods (Black & Wiliam, 1998b).

Since formative evaluation requires more student and teacher interaction it is, perhaps for logistical reasons, less commonly employed in an online environment (Wang, et al., 2006). While formative evaluation does not require face-to-face interaction it does require access to how the student is thinking about a topic, and this can sometimes be more difficult when teacher and student do not meet in a physical classroom. For example, mathematics courses conducted online are particularly vulnerable because of the technological constraints. Since most problems requiring formulas and equations need to be done by hand it is cumbersome for students to show their work digitally. Many students do not have scanners or want to use them to digitize their work. Without the digital version of the work it becomes unlikely that teachers can then look for problems in the thought processes behind the students' answers.

Because online instruction is the fastest growing part of education (Allen & Seaman, 2007) more students will be taking their mathematics classes at a distance. It is important to find out if this puts them at a disadvantage because currently there are not many convenient technological options to perform formative evaluation in an online mathematics class. If the proposed research can demonstrate that using formative evaluation that informs students about buggy algorithms, lack of structural knowledge, errors in logic, and misconceptions is effective, it would be worth considering how to overcome the technological barriers. The next section will look more closely at the types of errors that formative evaluation uncovers.

## Relevant Research Literature

To properly investigate the problem a careful review of the literature was undertaken to determine an appropriate course of action. Particular attention is paid to the types of errors students are likely to make while working mathematics problems. In addition, the practice of formative assessment is investigated. Special attention is paid to formative assessment as it applies to mathematics and in particular geometry.

### Types of Errors

The Program for International Student Assessment (PISA) released its results at the end of 2010 and it was revealed that 15 year olds in the U.S. ranked 25th among peers from 34 countries on the mathematics portion (PISA, 2011). Regarding the results Education Secretary Arne Duncan said, “The brutal fact here is there are many countries that are far ahead of us and improving more rapidly than we are. This should be a massive wake-up call to the entire country” (Hechinger, 2010, para. 4). What exactly we are supposed to be waking up from is not as clear, since learning mathematics is an extremely complex activity and there are many areas where things can “go wrong” for students when they attempt to solve problems. Of particular concern to the proposed study are four sources of errors that are different, but related. These are: buggy algorithms, misconceptions, errors in logic, and lack of structural knowledge. These are important because they are the four types of non random errors that are expected to be seen in the proposed study.

A buggy algorithm is an error in the logical steps and/or procedures that needs to be used to solve the problem (Li & Li, 2008). A misconception is the result of a student misunderstanding the basic concept of what is taking place (Li & Li, 2008). An error in logic is a breakdown of the deductive process where the reasoning for an answer was not provided the needed degree of support from the premises (Shaughnessy, 1985). Lack of structural knowledge is any missing knowledge that is a necessary condition for successful problem solving (Mayer, 1984). In addition, Mayer (1984) described four types of knowledge necessary to solve problems: linguistic and factual, schematic, algorithmic, and strategic.

It is perhaps easiest to give an example to illustrate the differences between the various errors. Take the 45-45-90 triangle where a student must solve for the length of the two missing legs.

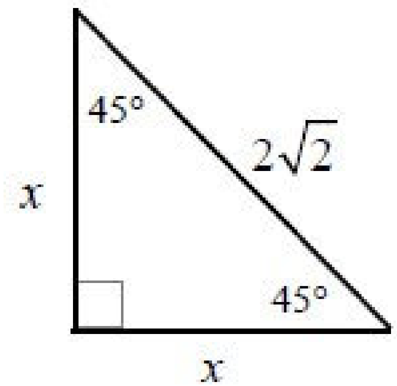



Figure 1.0

If students are given instruction on triangles where the examples are too often equilateral it is possible they could develop an understanding that all triangles must have three sides of the same length.  This would be an extreme example of a misconception of triangles, which would likely result in the wrong answer (incidentally many children have this misconception of triangles (van Hiele, 1999).   If students' attempt to solve the problem with the formula: Leg = Hypotenuse x √2 (reversing the leg and hypotenuse in the correct formula) then they are using a buggy algorithm, which will result in the wrong answer.  Lack of structural knowledge can be thought of as missing information that is required to do the work.  So, in this case if students have forgotten the formula for finding the legs of a 45-45-90 triangle, rather than misconstruing as in the buggy algorithm, they would be simply lacking the knowledge needed to solve the problem. Finally, if students are aware that sides of a triangle have a corresponding relationship to the opposite angles and yet their answers indicate that the legs are larger than the hypotenuse or determines the legs are both different sizes, then an error in logic has been committed.  That is, the answer is not supported by a known premise.  (In this case, the premise alone would not give the correct answer, but would tell them what they had was incorrect).  It should be noted that if a student simply supplied a teacher with a wrong answer, like found on a multiple choice test, it would be far more difficult to determine, which, if any, of these errors occurred.

While these errors are different in definition they also have aspects that overlap. For example, if a student is lacking structural knowledge about a key piece of information it can lead to a misconception (Mayer, 1984). If a student is aware that a hypotenuse is the largest side of a triangle yet uses the buggy algorithm, Leg = Hypotenuse x $\sqrt{2}$ it can also be seen as an error in logic since multiplying the hypotenuse by any number greater than one to find another side is a clear violation of a known premise. In another case, students might be able to deduce an answer from known facts, but do not because they already have a prior misconception about the conclusion and they ignore their intuition. All the errors listed will be of importance to the study, and particular attention will be given to errors of logic. Unlike other fields of study this type of error is particularly important to mathematics (Shaughnessy, 1985). To examine this further consider this problem and the typical responses.

> Suppose that we have a quadrilateral with both pairs of opposite sides congruent.
> Are the opposite sides parallel? Why? (Shaughnessy, 1985, p. 405).

When asking the above question high school geometry students at the end of the course the researchers found two types of responses: “Of course they are because I can’t draw them any other way,” and “Yes, because if the opposite sides are parallel, then they must be congruent” (Shaughnessy, 1985, p. 405). The first response depends completely on a visual representation and the second uses faulty *converse* logic. After a full year of geometry the visual response students were not able to approach the question as a logico-deductive one (Shaughnessy, 1985). While those students who used faulty converse reasoning do not understand basic logico-structural rules. Even though a statement like, "if an animal is a rabbit, then the animal is a fast runner" maybe true it does not automatically make the converse true, "if an animal is a fast runner, then the animal is a rabbit." Mathematical statements in which converse statements are true are special cases called *if and only if* statements.

A study by Burger (1982) reveals how common it is for students to confuse the role of *necessary* and *sufficient* conditions. (To understand the difference think if a number is evenly divisible by four it is sufficient to tell you the number is even, but it is necessary that the number is evenly divisible by two to be considered even). In the study, Burger had students try to guess a shape of a quadrilateral as clues are revealed one-by-one. The students were able to ask the researcher to stop giving clues once enough information has been supplied to determine what shape the geometric item was. What was found is that all students studied stopped too early and guessed the wrong shape. This is a result of considering the clues as necessary conditions rather than sufficient (Shaughnessy, 1985).

Several studies have demonstrated how vulnerable people are to making these errors in logic. For example, Oaksford and Chater (2001) found that when testing college students reasoning skills in a laboratory setting error rates can be as high as 96%. Also, Johnson-Laird and Wason (1977) studied people's reasoning processes on deductive tasks. They found that the subjects did not understand the difference between necessary and sufficient conditions. Subjects would also interchange the hypothesis and the conclusion exhibiting the same faulty converse reasoning as in Burger's study (Shaughnessy, 1985).

Clearly there is a reason to be mindful of these kinds of errors, especially with regards to mathematics. For geometry, in particular, this kind of knowledge would seem crucial.

Shaughnessy (1985) suggested that without it students could only think of mathematical concepts at a preconceptual stage. For example, without logico-structural knowledge it would be impossible to define a rectangle. For Shaughnessy (1985) the preconceptual student can only see a rectangle as a "bundle of properties" (p. 406) that are necessary conditions. It is only through logico-structural knowledge that a student can begin to refine the "bundle of properties" into "necessary and sufficient conditions for a definition of rectangle" (Shaughnessy, 1985, p. 407).

If we are to answer the wake-up call that Education Secretary Arne Duncan has asked for, we need a window into the thinking of how students are solving mathematical problems. Only when this is done can teachers begin to understand how errors are occurring and offer an appropriate remedy. The next section reviews the proposed method to uncover these errors that seems vital for teachers to be doing for online mathematics students.

**Formative Assessment - General**

The basic concept of formative assessment has been practiced by educators for centuries. The origins, however, of formative assessment as a formal theory can be traced back to Scriven's (1967) description of formative evaluation in education (Allal & Lopez, 2005). The seminal work in this field comes from Black and Wiliam (1998a). One citation index shows their paper has been cited over one thousand times. In the paper they state that the evidence "shows conclusively that formative assessment does improve learning." (Black & Wiliam 1998a, p. 61). With Black and Wiliam's review there has been an explosion of interest in formative assessment.

A careful examination of the formative assessment literature demonstrates the absence of agreed upon definitions. For example, Black and Wiliam (1998a) defined assessment as "all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged" (p. 17). These activities could include reading and writing exercises, lecture checks, group work, board work, class debates, etc. Sadler (1989), in contrast, defined formative assessment as, "a process used during instruction to provide feedback for adjustment of ongoing teaching and learning for the purposes of improving student achievement related to instructional objectives" (p. 120). Taras (2009) pointed out that Black and Wiliam have inconsistent processes of formative assessment listed in the same paper and only one is supported by their definition. The first process, which follows the definition, centers on the concept of classroom interaction where the teacher is the instigator of the process. The second, which does not correspond to the original definition, seems to be more of a "dramatic conceptual leap" (Taras, 2009). This process involves identifying a desired level of knowledge versus the actual level and then using that information to close the gap, which is more closely related to Sadler's definition.

Moreover, much of the assessment literature is devoted to making a clear separation between summative and formative assessment. Summative assessment, however, can be used with formative intentions (Bell & Cowie, 2000). For Black and William an assessment is only formative if it is used to provide feedback to the teacher and/or student. For them "feedback to any pupil should be about the particular qualities of his or her work, with advice on what he or she can do to improve, and should avoid comparisons with other pupils" (Black & Wiliam, 1998b, p. 142 ). Other researchers define feedback as information about the gap between actual and predetermined levels (Ramaprasad, 1983; Sadler, 1989). Usually it will be mentioned that the feedback must be used to modify the teaching and learning process. Technically speaking, however, an assessment that assigns only a grade (i.e., “A” or “B”) does provide feedback and could be used in the future to adapt the teaching and learning process. This would be classified as formative assessment under many researchers’ definitions—if even in a limited way.

For Taras (2009) breaking formative and summative assessment into two divisions is a false dichotomy. She pointed out that for formative assessment to take place a judgment of the quality of work must have happened and this judgment is a summative assessment. Sadler also believed that the initial assessment must be a summative assessment and by not naming the initial step of summative assessment, the whole process of formative assessment seems to have become confused by researchers (Taras, 2009 ).

Regardless of how an assessment is designed or packaged it is the use of the results that determines whether an assessment is formative or summative. The literature is filled with ambiguous uses of terms. For example, one researcher used a summative assessment, but incorporated formative assessment by providing both quantitative and qualitative feedback about the results. That is, feedback both in the form of written comments was given to the students as well as a grade on the test. This was termed “formative summative assessment” (Wininger, 2005). This exemplifies complications that arise when one defines assessment by usage. As Dunn and Mukvenon (2009) wrote, “an assessment is an assessment, and the manner in which an assessment is evaluated and used is a related but separate issue” (p. 2).

By separating assessment from evaluation it brings the field back to its roots when Scriven (1967) originally defined the term as “formative evaluation.” Some researchers have traced the complications of the field back to when Bloom (1969) transferred the term "formative" from evaluation to assessment (Dunn & Mulvenon, 2009). An assessment can be designed for a certain purpose, but teachers can use the data gathered in any way they see fit. It is how the information is being used (evaluated) that is important. For the purposed research project this evaluation must be fed back to the student with the requirement that they act on this information to make improvements. In the next section priority is given to operationalizing assessment as something unique from evaluation and this begins to turn the focus to usage of assessment making the research propose more clear.

**Formative Assessment - Definitions**

Because of the lack of commonality in the definitions of the terminology relating to assessment it is prudent to define carefully how the proposed study will be utilizing terms. In this case, the definitions given are the ones that best fit the claims of how the influence on the dependent variable is derived.

Some researchers, in a well-meaning attempt to meld summative assessment to formative assessment, (Taras, 2009; Salder, 1989) seem to have caused even more confusion in the field. Clearly there are different flavors of assessment and how a teacher uses the information from the assessment can differ regardless of the intended purpose of the assessment. Separating assessment from evaluation alleviates many problems present in the literature.

Since it adds clarity to distinguish assessment and evaluation this will start by defining the various components of assessment and then move to evaluation. An effective definition of assessment is something that refers to judgments of performance (OECD, 2005). Summative assessments then are those assessments designed to determine a students' academic development after a set unit of material (i.e., assessment of learning) (Stiggins, 2002). Formative assessments are assessments designed to monitor student progress during the learning process (i.e., assessment for learning) (Chappuis & Stiggins, 2002).

On the other side, evaluation is defined as utilizing assessment-based data for a specific purpose. Summative evaluation is the evaluation of assessment based data for the purposes of assessing academic progress at the end of specified time period (e.g., a unit of material or an entire school year) for the purposes of establishing a student's academic standing relative to some established criterion (Dunn & Mulvenon, 2009). Finally, formative evaluation is the evaluation of assessment-based evidence for the purposes of providing feedback to and informing teachers, students, and educational stakeholders about the teaching and learning process (Dunn & Mulvenon, 2009).

These definitions not only help operationalize the terms, but they also help clarify many of the ambiguities found in the terms throughout the literature. While it is common in the literature to refer to the practice as formative assessment it would be more specific to call it formative evaluation.

The above definition of formative evaluation qualifies for my proposed study, but it would be best to be even more explicit since this would be the independent variable of the proposed experiment. For this research, formative evaluation is defined as the evaluation of assessment-based evidence for the purposes of providing feedback to students regarding the rationale of their responses in an attempt to correct buggy algorithms, lack of structural knowledge, errors in logic, and misconceptions in their thought processes.

**Formative Assessment – A Critical Review**

For many it is a given that formative assessment improves student performance. After all, not only do Black and Wiliam (1998a) state that the evidence is conclusive, but that gains under its use are "amongst the largest ever reported" (p. 61). Black and Wiliam's conclusion is drawn after reading 700 journal articles on assessment and narrowing the list down to 250 as being relevant. From this they whittle it down to, what are deemed, eight top-notched research studies that give them their evidence that formative assessment does work. Some researchers, however, have found issues with all eight studies (Dunn & Mulvenon, 2009).

The biggest problem with the studies is issues with generalizability. Black and Wiliam's conclusion most strongly relies on the meta-analysis of Fuch and Fuch (1986). While this had 3,835 participants, from the various studies, 83 percent were handicapped. This is because the review was specifically focused on special education. Although there was an average effect size of 0.63 for the non handicapped participants it seems inappropriate to generalize this to all students given the slant of the sample (Dunn & Mulvenon, 2009).

Several of the studies used by Black and Wiliams do not account for teacher effects. For example, one study had a massive sample of 7,000 students viewed over eighteen years (Whiting et al., 1995). Only one teacher, however, was used in the study. While the teacher did use formative assessment and was directly compared to a teacher who did not use formative assessment it is difficult to ignore the potential confounding variables.

Dunn and Mulvenon (2009) reviewed nine more recent articles on formative assessment. While the articles do lend more evidence to formative assessment they also suffer from methodological issues. For example, one promising study gave impressive results, but is difficult to generalize due to the sample size of four. In addition, while Black and Wiliam (1998) reported effect sizes of 0.70, the effect sizes of more recent studies tend to be smaller (Buchanan, 2000).

While it is a popular notion in education that formative assessment has already been proven to work, clearly more research needs to be done. The topic has both great potential and vulnerability due to the lack of strong empirical evidence. There are many questions that can be explored further. For example, which students benefit the most from formative assessment? Under what conditions does formative assessment appear most effective? Are the positive effects of formative assessment independent of who is teaching the course? Which subjects are given the greatest gain? What learning environments does formative assessment work well with? For this proposed study the focus will be on how formative evaluation affects online collegiate community college mathematics, which is explored in the next section.

**Formative Assessment - Mathematics**

In mathematics it is important for the teacher to understand how students solve problems (Wiliam, 1999). This knowledge lends insight into the thought process of students and makes it easier for teachers to overcome their students' conceptual difficulties. This elevates formative assessment to the diagnostic level. It has been shown that, in particular, formative assessment can give information about a student's performance, thinking, and learning potential (Ginsburg, 2009).

When researcher Herbert Ginsburg (2009) started with the question, "How can the new psychology of mathematical thinking be used to improve mathematics education?" (p. 109). His conclusion was, "to inform, indeed transform, the process of 'formative assessment'" (Ginsburg, 2009, p. 109). For Ginsburg (2009) using formative assessment in mathematics starts with an understanding of the three methods that it can be employed by: observation, clinical interview, and test.

The observation method has a long history in psychology (Piaget, 1952). Piaget actively observed his children to develop his stages of development theory. Ginsburg (2009) states, "Much can be learned from keen observation. We see that a child playing with two blocks is exploring ideas of symmetry and pattern" (p. 112). It is rare that observation is pure and typically involves some interaction with the subjects being observed (Ginsburg, 2009). Also, it is not as simply as it may seem. Generally it involves a great deal of thinking about what to be on the lookout for. While this may also be fruitful for adults it seems impractical as an approach to formative assessment in an online class.

Because it is inconvenient to wait for students to do something important to observe the clinical interview is another approach. In this case, the interviewer develops a task and asks the subject to complete it. This is useful because the interviewer can stop the subject and ask questions along the way. For example, "what makes you decide to start at that point?" or "how did you do that?" For Ginsburg (2009) this is the most powerful formative assessment technique and offers "deep insight into thinking" (p. 115). Typically the interviewer would start with an assumption and give a task while asking questions in an attempt to verify the hypothesis. This can be difficult because it requires the interviewer to make interpretations and act quickly on those interpretations (Ginsburg, 2009). Because this is a focused method that still offers a teacher a lot of flexibility it has large benefits. Again, this would be hard to use for online teaching, although it could be a useful idea for online office hours where synchronous communication would make it easier to manage interviews. In this case, the teacher could ask some questions to gauge where the student is in the learning process and offer suggestions on remediation.

While the other two approaches are used far less in education the most common formative assessment method is the use of tests or tasks. These have the benefit of getting right to the important questions a teacher wants to ask. Ginsburg (2009) believes much can be learned from tests and it is not just checking for right and wrong answers, but "strategies of solution" (p. 113). Ginsburg (2009) writes that the teacher "may observe what [a student] writes on paper to solve a problem, thus obtaining information about the strategy" (p. 113). This immediately points to one of the key problems of teaching mathematics online mentioned in the problem statement section; that is, a teacher is not privy to what the student "writes on paper to solve a problem."

At the same time Ginsburg points out the drawbacks of tests. He tells how Piaget (1952) reported that working as an administrator of standardized intelligence test he found responses to be of less interest than the cognitive process that produced them, but could not be inspected because of the technique of standardized test administration (Ginsburg, 2009). The idea is that using formative assessment in testing is useful, not for the answers, but instead the thought process behind the answers. There are, however, few options for this in online testing of mathematics. It seems safe to say that Piaget would be frustrated with the state of online mathematics education today.

As noted by Ginsburg (2009), "formative assessment must rest on a foundation of mathematical understanding" (p. 117). When students' assimilate new information into what they already know, misconceptions can result. Often the student takes a principle that might be useful in certain contexts and generalizes it in an inappropriate manner. Knowing this requires an insight into how the student is deriving an answer. Given Ginsburg's three methods of using formative assessment in mathematics (observation, clinical interview, and test) the problem is that with current technology none of these approaches are very effective in an online environment. The first two methods have the problem of distance, whereas, the last is because students will rarely take the time to capture their work with equations digitally.

Geometry, however, can often allow students to express their work in words without the use of equations. By simply typing in how the answer was derived, teachers can see the students' thought processes and offer formative evaluation. Using geometry for the proposed study will be looked at closer in the next section.

**Geometry**

While geometry is often relegated to high school students, since the time of Plato it has been considered one the best methods to develop logical thinking (Hogben, 1993). Geometry has an advantage over other subjects because it can easily be taught from early childhood through graduate school. By noticing difficulties that students had in geometry researchers develop a theory that involves specific levels of thinking in geometry. Students start at the rudimentary level of merely recognizing figures and must pass through each level as they progress to being able to construct formal proofs. These are called the van Hiele levels of reasoning in geometry and consist of five levels of understanding. These include: visualization, analysis, abstraction, deduction, and rigor (Burger & Shaughnessy 1984).

The initial stage (level 0) is where students reason about basic geometric shapes. This focuses on learning vocabulary, recognizing shapes and reproducing shapes. At level 1 (analysis) the student begins to discern characteristics of the figures. A student will recognize that the figures have parts. At level 2 (abstraction) the student can begin to recognize interrelationships between figures. Here the student can deduce properties and logically order them and distinguish between necessary and sufficient. Level 3 (deduction) is when students can begin to construct formal proofs and they can grasp the underlying logical system and work with the theorems and axioms. Level 4 (rigor) is where a student can begin to work with different geometries (non Euclidean) and compare how the two systems operate (Burger & Shaughnessy 1984; Crowley, 1987).

For the purposed study, content covered will be geometry at the deduction level. Although the majority of high school geometry classes are taught at this range, Mayberry (1983) has found that many high school students never reach the level of formal deduction, a conclusion shared by Usiskin (1982) (Burger & Shaughnessy 1984; Crowley, 1987). This suggests that deduction level is a good place to start since it should be challenging, yet not too overwhelming for the community college students that would be included in the study.

Geometry is also useful to use as the course content because the study purposes to use formative evaluation to look for errors in logic, among other things, and geometry has a strong emphasis on using logic. Finally, because the proposed experiment depends on students receiving formative evaluation it requires that evaluators have a window into the thinking of students. Since getting written work digitized is one of the big constraints in making this happen, geometry is ideal because it offers a convenient solution. Unlike algebra, where work often requires using equations and special notation, in geometry the work can usually be written out in words. The idea is to simplify the process as much as possible by requiring students to write out the solution without the use of drawings. This will allow the evaluators to look for buggy algorithms, lack of structural knowledge, errors in logic, and misconceptions in the students' thought processes simply by reading how the problem was solved. It is vital to see how students are "constructing" the content they are given and this is something that will be discussed more in the next section.

## Theoretical/Conceptual Framework

Constructivism and motivational theories lie beneath formative assessment. Understanding these theories is crucial to explain how formative assessment works and why. This understanding is also important because the theoretical basis for formative assessment ties together separate elements of effective practice and helps to clarify them and see how they fit together.

### Constructivism

The early research on mathematics education viewed student errors as problems that needed to be avoided, and misconceptions as something that needed to be replaced with accurate information (Even & Tirosh, 2008). This stems from a behaviorist view of learning. With this approach the learner is passive recipient responding to environmental stimuli. The student is viewed as a clean slate (i.e. *tabula rasa*) and behavior is shaped through positive reinforcement or negative reinforcement. For education purposes the underlying assumption is that knowledge can be transferred intact from one person to another and any current knowledge a student has is irrelevant to the learning process (Olivier, 1989; Mamba, 2011).

In contrast to the behaviorist perspective, a constructivist perspective on learning (Smith, diSessa, & Roschelle, 1993) assumes humans generate knowledge and meaning from an interaction between their experiences and their ideas. Here the current knowledge a student has will help determine what the student will learn. With behaviorism it is solely the experience that makes the associations that determine what is learned. With constructivism it is the interaction between cognitive structures and experiences.

An important concept in regards to constructivism is mental schemas. For the purposes of the proposed research schemas will be focused on from a Piagetian perspective and be defined as a cognitive framework organized around a theme that helps interpret information (Piaget 1952). In particular, the important theme will be logical-mathematical schemes. As Derry (1996) describe them, “These schemas have come to represent the ‘big ideas’ underlying mathematics understanding that student must construct for themselves” (p. 167).

For the behaviorist learning simply requires making new associates from the environment, but for the constructivist learning often requires assimilating to existing schemas. Misconceptions are important to this view because they exist in schemas and interact with newly assimilated information. These misconceptions can twist and distort information generating errors. Because of the nature of learning, according to this view, misconceptions are considered inevitable and can be used as an important source of information about the learning process. This means teachers should regard errors as a clue for uncovering what students already know and how they have constructed such knowledge (Borasi, 1996; Maba, 2011).

From the constructivist perspective learners make sense of new information when it is incorporated in mental schemas (Smith, diSessa, & Roschelle, 1993). Often a teacher's job is to help students' organize how well their schemas are connected (Donovan & Bransford, 2005). While teachers can assist in the process they cannot change students' schemas directly. J Smith states, "Change has to happen internally and organically as students' try to make sense of new experiences and new ideas" (personal communication, September 1, 2011). This is done by having the student interact with where they currently are and where they need to be. This interaction can come from teachers, peers, or even careful diagnostic feedback (Reveles, Kelly, & Durán, 2007). Interaction works best if it is close to the boundary of what a student knows and does not know, called the zone of proximal development (ZPD) (Vygotsky, 1978). For Vygotsky (1978) learning was a social process that required interacting with others.

Vygotsky's theory distinguishes two levels of development. The first is the current level of the individual. The second is the level of potential development. This potential development is the level that the student is capable of reaching with the assistance of a teacher or peers (Heritage, 2010). To guide in this process it is important to build scaffolding for the student (Belland et al., 2008). For example, a teacher may ask "leading or probing questions to elaborate the knowledge the learner already possesses, or providing feedback that assists the learner to take steps to move forward through the ZPD" (Heritage, 2010, p. 8). As the learner becomes more competent in the subject area the scaffolding is gradually reduced until the learner is able to function independently (Heritage, 2010; Tharp & Gallimore, 1988).

The connection to formative assessment is clear. Formative assessment enables teachers and students to consistently work in the zone of proximal development (Heritage, 2010). By its nature formative assessment involves an interaction or a dialogue between teacher and student and it is this interaction that helps to build what Vygotsky (1978) calls “maturing factors” (p. 86). By carefully utilizing formative assessment teachers can monitor errors that crop up in the students assessments and use that information to support learning (Heritage, 2010). Finding these errors by using formative evaluation will allow the evaluators of the purposed study to feed the information back to the students and give the students the opportunity to confront their buggy algorithms, lack of structural knowledge, errors in logic, and misconceptions. When students are allowed to do this the belief is that it empowers them and gives them increased confidence in their ability. How to feed the information back to the students and increase their confidence will be discussed in the next section.

**Feedback**

For Sadler (1989) feedback is the crucial factor to aid learning. It is usually defined by giving information about how something is being done. Sadler (1989) instead prefers a systems perspective of feedback and uses the definition “information about the gap between the actual level and the reference level of a system parameter that is used to alter the gap in some way” (p. 4). With this approach feedback loops are important and formative assessment is the key to closing the gap. From a constructivist perspective J. Smith states, “feedback is only ‘real’ to the learner if he/she can ‘take it in’ in some way. Feedback offered is not necessarily feedback received” (personal communication, September 1, 2011).

To create these feedback loops requires students to practice in a supportive environment that usually includes a teacher who understands the skills that are needed for improvement and how to correct a poor performance. For Sadler’s model the feedback is used both by teachers to make programmatic decisions and also by students to monitor their strengths and weaknesses. The key is that it is only feedback "*when it is used to alter the gap*" and is useless if the "information is simply recorded" (Sadler, 1989, p. 121).

Effective feedback focuses on the task and provides the student with suggestions, hints, or cues; feedback in the form of praise is problematic (Kluger & DeNisi, 1996). For the feedback to be successful it must be prompt and accurate. The closer the feedback comes to the assessment the more effective it is for student achievement (Waterman, 2010). Wiliam and Leahy (2007) advise three *time scales cycles* for feedback: short, medium, and long. Short is defined as between five seconds and one hour. Medium is between one day and two weeks and long is between four weeks and one year.

Short time scale feedback is usually done to help a teacher decided whether to proceed to a new topic or explain the content again. Medium time scale feedback is better used to determine if an entire class needs to be retaught or to take aside those students that have not learned to reteach them in a special session. Long time scale feedback is less common and is used to determine if students are needed to be regrouped for reteaching learning objectives (Wiliam & Leahy, 2007; Waterman, 2010).

Further evidence points to the role of effective feedback in the learning process. In a review of 196 studies describing nearly 7,000 effects it was reported that feedback had an average effect size of 0.79 (Hattie & Timperley, 2007; Heritage, 2010). If realized, this size effect would be one of the largest in educational interventions and would have a profound impact on a student's success. Sadler, however, was concerned with the puzzling results that some studies showed that even when teachers provide feedback there was no improvement or even worse feedback was shown to hurt improvement (Wiliam, 1999). One answer to the conflicting findings is explored in the next section on motivation.

**Motivation**

Attribution theory is a good place to start when looking at what motivates students academically. It focuses on how people attribute the causes of events and how these judgments influence internal perceptions. It is tied with the concept of self-efficacy, which are the students' beliefs in their own competence. The students' self-efficacy will play a role in how they will interpret success or failure.

In mathematics the self-efficacy of student's can be defined as their internal perceptions of the potential for their success (Wolters & Rosenthal, 2000). It has been shown that students with higher levels of self-efficacy set higher goals, apply more effort, and persist longer on difficult tasks (Tanner & Jones, 2003). Students with this high self-efficacy do not believe that mathematical ability is fixed and they attribute their success or failure to internal factors (Black, 1998). Black (1998) worried that when a student repeatedly gets low grades it may cause "a shared belief between them and their teacher that they are just not clever enough" (p. 43). Thus, low self-efficacy can be the unintended result of summative assessment (Tanner & Jones, 2003). This can create a vicious circle where a student fails, which lowers self-efficacy that results in more failure.

A study by Butler (1988) demonstrated that giving grades does boost students' interest, but only when the grades are high. As the grades dropped interest plummeted. In another study, Butler (1987) broke 200 students into four groups and individuals of each group were given one of four kinds of feedback on a lesson: comments, grades, praise, and no feedback at all. On the second lesson only those who received comments showed improvement over the first lesson. At this time students were also given a questionnaire to see if they attributed their success or failure to ego-involvement or task involvement. Those who were given comments had high levels of task-involvement but their ego-involvement was the same as those that received no feedback at all (Wiliam, 1999). While the two groups that received praise or grades both had high ego-involvement and low task involvement. In addition, they demonstrated no improvement in performance. Is having high ego-involvement a good thing if it does not improve performance?

In a review of 131 studies on feedback it was found that, on average, feedback did improve performance (Wiliam, 1999). A closer look shows a significant difference between studies. A whopping 40% of the studies showed giving feedback actually had a negative effect on performance (Wiliam, 1999). On further investigation it was found that those studies where feedback hurt performance the feedback was focused on self-esteem or self-image (as in the case of grades or praise) (Wiliam, 1999). This suggests the importance of the kind of feedback used. In particular formative feedback is recommended (Black & Wiliam, 1998b). Feedback is considered formative “only if the information fed back to the learner is used by the learner in improving performance” (Wiliam, 1999, p. 8).

Feedback promises to be a powerful performance improver and motivator. Looking more closely at research, however, demonstrates that certain kinds of feedback are more promising than others. Because use of feedback is key in formative evaluation, and in particular in this proposed study, careful attention will be paid to how feedback is managed. Specifically formative feedback that focuses strictly on the work will be utilized in the evaluation process. The purpose, questions, and design of the proposed study will be covered in the preceding sections.

## Purpose Statement

The purpose of this study is to assess the effectiveness of formative evaluation as compared to traditional assessment methods with regards to college students' learning and interest in mathematics in an online environment. Specifically, it will be looking for some significant difference in performance, interest, and transfer of learning between the two groups.

## Research Questions

RQ1) Does using formative evaluation improve *performance* in course content for students studying mathematics online relative to only summative evaluation?

RQ2) Does using formative evaluation increase *interest* in course material for students studying mathematics online relative to only summative evaluation?

RQ3) Does using formative evaluation increase *transfer of learning* for students studying mathematics online relative to only summative evaluation?

## Hypotheses

The hypothesis is that using formative evaluation will result in a significant increase in learning gains, interest in the subject matter, and transfer of learning when utilized by students learning mathematics online.  The rationale is that when formative evaluation is used it motivates students to explore and identify problems in their own thinking.  The constructivist approach theorizes that students will generate knowledge from an interaction between their experiences and their ideas.  This is thought to allow the students to learn the material on a deeper level and will be seen with increased learning gains.  In addition, if students understand the material on a deeper level it believed that it will show increase transfer of learning.  When students are given the chance to interact with their misconceptions and errors it is believed it will increase their self-efficacy, or their confidence for success.  If true, this should increase interest levels in the topic.

## The Design--Methods and Procedures

The proposed study will use an experiment where performance, interest, and transfer of learning will be measured between two groups in which students will be randomly assigned with some receiving only summative evaluation and some receiving both formative and summative evaluation.  To measure performance a two-group posttest-only experiment will be implemented using a one-way analysis of variance.  To verify the equivalence of groups from the random assignment student grade point averages will be collected and compared between the two groups.  A t-test will be performed between the mean GPA's of the two groups to determine if the groups are equal in terms of their starting GPA.

Interest will be measure in two ways. First a t-test between the number of mouse clicks of the two groups within the online site. Each individual's movement in the site will automatically be tracked by the *summary of usage* feature built into the course management system. In addition, a questionnaire about the students' interest in the subject will be administered (see appendix). A t-test will be used to compare questionnaire results between groups. Finally, an item analysis of the posttest will look at the logical matrices, not covered in the content, to analyze any difference in transfer of learning. Logical matrices are designed to be simple to use and do not require special instruction to complete. These matrices test how well a subject can use logical skills in a visual environment, which has a strong correlation to geometry (Pind, Gunnarsdottir, & Johannesson, 2003; Koenig, Frey, & Detterman, 2007).

The study will be conducted in an online environment. The goal is to have between 80 and 150 college students. Subjects will be recruited from a population of students who have taken or are currently taking a course at Dallas County Community College District and many of the students would be participating to receive extra credit. Students will then be randomly assigned to one of two groups. Two teachers will be trained to look for buggy algorithms, misconceptions, errors in logic, and lack of structural knowledge in the students' work of the experimental group to provide appropriate formative evaluation when finding these errors.

**Site**

Students will be drawn from the Dallas County Community College District (DCCCD) which is the largest undergraduate institution in the state of Texas. It includes seven colleges — Brookhaven, Cedar Valley, Eastfield, El Centro, Mountain View, North Lake and Richland. The student population is diverse: 24.3 percent Hispanic, 23.9 percent African-American, 8.4 percent Asian, 39.9 percent Anglo and 3.4 percent all others combined (Who We Are, 2011). Most students are attending the institution to earn an associate's degree or the first two years of a bachelor's degree. It is estimated that up to 500 students could receive a request to participate.

**Sampling**

The population of this study would be undergraduate students studying mathematics at community colleges in the United States. For convenience, however, the sample will be drawn from DCCCD colleges only. In addition, while the population of the district is diverse it will not be a random sample. That is, not all students taking a mathematics class will have an equal chance of being represented. Since each semester there are many mathematics classes offered in the district, it cannot be guaranteed that students in all classes will have an equal chance to participate. This is because it will greatly depend on teachers publicizing the need for student subjects and in some classes it may not be done at all and in others teachers may offer extra credit to participate. The goal is to have between 80 and 150 college students. Because the duration of the experiment is short having a large number of students will make it easier to see an effect on the experimental group.

**Recruitment of Instructors**

A big part of my responsibility in the study will be giving students summative and formative evaluation, but due to the expected size of the study two additional instructors will be utilized. To find these instructors an e-mail will be sent out to faculty members asking for volunteers. Any interested teachers will first be required to take the geometry modules assessment that the students will be given, and they will need to receive 100 percent on the assessment to be eligible to participate. In addition, they will be asked to participate in a training session (see the instructor training section below). In the training session the instructor will need to demonstrate an alignment with my feedback style to qualify. This will be determined after the training session with an assessment that will be administered to see if the instructors can correctly identify the types of error committed by students, and if the feedback given for those errors is similar to my own. Finally, instructors will need to have at least two hours of free time a day for the duration of the experiment to assist in the evaluation process.

## Instructor Training

Instructors interested in providing evaluation will first need to demonstrate proficiency in the content matter. This means they will need to have a perfect score on the assessment modules that will be given to the students. After this instructors will attend two sessions. The first session explains the overall purpose of the experiment and will cover the types of non random errors that will be focused on in this study. Those errors being: buggy algorithms, misconceptions, errors in logic, and lack of structural knowledge. The second session will demonstrate the kind of formative feedback to be given depending on the type of error seen. At the end of the second session instructors will be required to demonstrate an alignment to my formative evaluation approach. This will be done by an assessment to see if the instructors can correctly identify the types of error committed by students and if the feedback given for those errors is similar to my own.

**Recruitment of Students**

By working with teachers across the Dallas County Community College District the goal is to recruit as many mathematics students as possible. To recruit students several methods will be employed. One will be meeting with mathematics faculty members to explain what is needed and request that they offer the opportunity for their students to join the research experiment. Also, a flyer will be sent electronically to mathematics faculty members throughout the district with a request to distribute to students. Finally, there will be some direct marketing through an e-mail list of past students that have taken my classes.

An emphasis will be on recruiting students from a wide variety of mathematics class levels. After the students have been selected they will be randomly assigned into either the experimental group or the control group. The mean score of the GPA of the two groups will be determined and a t-test will verify that there is not a significant difference between the two groups. Although the population of the study is undergraduate students, this study could potentially generalize to other level mathematics courses online.

**Groups**

The groups will be divided into a control group and an experimental group through random assignment instead of matching. The two groups will be placed in an online course management system—Blackboard version 9.1. The course will be available over two-weeks (although most should complete in one-week or less) and two modules will be covered. There will not be a set time that students need to be available and they can log on at their convenience and watch the geometry content using a multimedia presentation. After watching the content from the module they will take a quiz over the material that was seen in the multimedia presentation. The control group will only be given a grade on the quiz and the correct answers. The experimental group will not be given either a grade or the answers, but will only receive formative evaluation and the requirement of going back to the quiz and correcting any errors. At the conclusion, each group will take the same post-test and be given an identical questionnaire.

The material covered is not likely to have any relation to what the students are studying in their classes. That is, they will be learning mathematical concepts just for the experiment.

## Materials

For this experiment several geometry text books will be used to build the geometry assessments. The assessments will be supported by learning modules. These modules will be multimedia presentations built with the Articulate software. Also, an assessment for the instructors will be created to verify their understanding of the evaluation process and to ensure their feedback is aligned with my own. Finally, an informed consent form will be developed to give to all participants, so they understand their rights in the experiment.

**Geometry Modules**

There will be two geometry modules. The first will cover triangles and the second will include circles. In the triangle module the following topics will be explained: altitude, sides and angles, perimeter of a triangle, area of a triangle, right triangles, Pythagorean Theorem, 45-45-90 triangles, and 30-60-90 triangles. In the circle module the following topics will be covered: diameter, radius, central angle, circumference and arc length, area of a circle, and sector. The modules will be multimedia presentation done with Articulate software. Students will be able to view them as often as needed before taking the assessment. These assessments are covered in more detail in the next section.

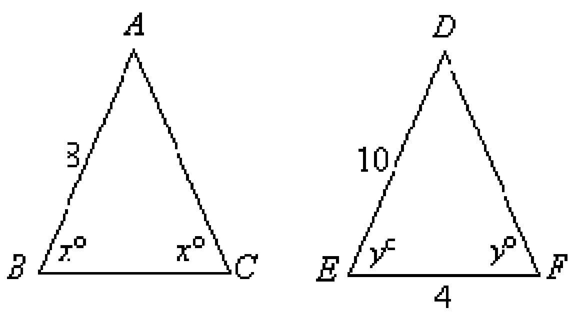
**Geometry Assessments**

After reviewing each module the students will take an assessment. A key requirement is for the instructors to be able to see the work of the students in the experimental group in order to give formative evaluation. To prevent introducing a confounding variable both the control group and the experimental group will type in how the answers are being derived. In other words, both groups will be showing their work, but instructors will only look at answers of the control group. To keep the student from having to use cumbersome mathematical symbols they will be shown a brief list of keyboard alternatives:

| Process | Mathematical Symbol | Keyboard Symbol | Example |
|---|---|---|---|
| Division | ÷ | / | 6/3=2 |

| | | | |
|---|---|---|---|
| Pi | π | pi | Area=64pi |
| Square root | √ | Sqrt( ) | Sqrt(4)=2 |
| Exponent | $4^2$ | 4^2 | 4^2=16 |
| Angle | ∠ | < | <BAC=60 degrees |

An example problem from module one could look like this:




Note: Figures not drawn to scale

In the figures above, x = 60. How much more is the perimeter of triangle ABC compared with the triangle DEF.

A text box would be provided for the students to show their work. An ideal student response for this question would look like this:

Step 1: If *x* = 60 degrees, than triangle *ABC* is equilateral and all sides are equal.

Since we know one side is equal to 8 all the sides are equal to 8.

Perimeter of triangle *ABC* = 8 + 8 + 8 = 24.

Step 2: Triangle *DEF* has two equal angles, and therefore it is an isosceles triangle.

The two equal sides will be opposite the equal angles.

So, the length of *DF* equals the length of *DE* = 10.

Perimeter of triangle *DEF* = 10 + 10 + 4 = 24.

Step 3: Subtract the two perimeters.

24 – 24 = 0

("Geometry Questions," n.d.).

By having students show their work it gives the evaluators insight into the thinking about the problem. The next section will cover how the feedback will be utilized in the study.

**Geometry Assessment Feedback**

The subjects in both the experimental and control groups will type in the steps they are using to solve their problems. The instructors will then give feedback to all the students. For the control group feedback will only be acknowledgment of a right or wrong answer and if incorrect they will be given the correct answer. This kind of feedback mimics most online grading systems. With the experimental group formative feedback will be given. In the experimental group the correct answers will all be checked to verify that how the students derived the answer is legitimate. For incorrect answers instructors will look for the type of error that occurred and supply feedback. The correct answer will not be given and it will be expected that the student goes back and tries the problem again. Formative feedback would last for two rounds. This means that if the student gets an answer wrong on the first attempt formative feedback would be given. If the student gets it wrong on the second attempt again formative feedback would be given. On the third attempt the student would receive the same feedback as the control group (summative feedback) and that would end the student working on the problem.

Here is an example of how the feedback could work:

1. If the area of a circle is 64π, then the circumference of the circle is…

Student Work:

Step 1: The area = pi r^2

Step 2: pi r^2 = 64pi, so r =32

Step 3: Circumference = 2 pi r

Step 4: 2 pi (32) = 64 pi

Since we know that the radius is raised to the second power we need to take the square root of 64 to find r, but in this case it appears the student is dividing 64 by two and determining r=32. Since the student listed the correct formulas for both area and circumference of a circle the buggy algorithms is not an issue. The reason the student is ending up with 32 could be from a misconceptions of how square roots work or from a lack of structural knowledge that taking a square root is needed. In addition, an error in logic could also be present assuming the student knows that the area of a circle would always be larger than its circumference. All these concerns could be addressed. Appropriate feedback would look like this:

On question 1 good job with using the correct formulas. There was only one problem and that was when you solved: pi r^2 = 64pi. In this case pi is on both sides of the equation and cancels out. We are left with r^2=64. What is needed to solve this is to take the square root of both sides. That is because finding a square root is the inverse operation of squaring that number.

r^2=16 is solved by

Sqrt(r^2) = Sqrt(16)

r = 4

Remember, the square of a number is that number times itself. For example:

Sqrt (9) = Sqrt (3 * 3) = 3

Sqrt(25) = Sqrt (5 * 5) = 5

Also, be aware that the area of a circle is always larger than the circumference. For example, think of how many jellybeans that could be place around a hula hoop compared to how many could be put inside. Please rework the problem and resubmit.

There are three significant elements to the feedback. First, there is an attempt to address the specific errors that are believed to be contributing to an incorrect answer. Second, the correct answer is never given. (In the case above the student is not even given the square root of 64). Third, the student is given a requirement to use the feedback given and go back and rework the problem.

**Instrumentation**

In the proposed study four instruments will be used: a survey, a post-test, a questionnaire, and a summary of usage in the online environment. The survey will ask for the students' GPAs. The post-test will include geometry problems from the content that was previously covered. In addition, there will be questions using logical matrices to test transfer of learning. The scores from the post-test will be used to determine if there was a significant difference between the control and experimental groups. A questionnaire (see appendix) will be used to determine how much interest a student has in the material covered. And to further determine interest the summary of usage feature of the Blackboard course management system will determine how thoroughly the site was explored by individual students by tracking mouse clicks.

**Data Collection**

Data will be collected from the four instruments mentioned above: a survey, a post-test, a questionnaire, and a summary of usage. Student GPA data will collected from the survey at the beginning of the experiment to ascertain how equivalent the two groups are.  The rest of the data will be taken at the conclusion of the experiment.  All this information will be captured digitally in the Blackboard course management system.

**Data Analysis**

The mean GPA will be determined for each group.  A t-test will be used to see if there is any statistical significance between the two GPAs.  To measure performance a two-group posttest-only will be implemented using a one-way analysis of variance.  To measure interest both a t-test between number of mouse clicks within the online site and a basic analysis of covariance design will be used to compare questionnaire results with pre- and post-test as well as between groups.  Finally, an item analysis of the performance test will look at specific questions to analyze any difference in transfer of learning.

## Time Line

October 2011: Begin work to submit IRB

November 2011: Develop online course materials

January 2012 - February 2012: Work with faculty to help with volunteer subjects

February 2012 - Market study to potential volunteers

March 2012: Administer course content

March 2012: Data analysis

April 2012: Practicum draft report

May 2012: Final writing of practicum report

## Limitations

There are two significant limitations of the proposed study. The first is the fact it is using a convenience sample which is a threat to external validity. The subjects will not be drawn from community colleges across the nation, but from one district in Texas. Furthermore, the sample will not be randomly drawn from all registered students in a mathematics class. It will be dependent on volunteers from various classes. The district, however, has a diverse population both of ethnicities and socio-economic backgrounds. In addition, effort will be made to recruit from a variety mathematics classes and students with different ability levels. Once the subjects have been recruited they will be randomly assigned to one of two groups.

The other potential limitation is a threat to internal validity. Because students have demonstrated a high dropout rate in online classes there is the potential to lose a considerable amount of subjects. Also, the course will last two weeks taking approximately two to three hours to complete. My observation is that this is a significant amount of time for the students to commit and increases the likelihood that subjects will withdraw before completing the experiment. Careful attention will be paid to tracking usage patterns and monitoring who stays and who leaves, as well as what they do when they are there.

## Significance & Implications

This study will extend the existing research on formative assessment in online learning environments by providing new empirical evidence. After an exhaustive search, no research regarding the use of formative assessment in an online environment for mathematics has been uncovered.

If the hypothesis is confirmed it may require rethinking how students are assessed and evaluated online with regards to mathematics courses. Currently, online course management systems do not have a specific method that allows teachers to observe students' work when it comes to mathematics. This means any tests or quizzes that are done online would not give teachers access to the thought process behind the answers given. If the hypothesis is not confirmed then there may not be a need to overcome technological constraints of using formative assessment in an online environment. This knowledge will be useful to teachers that are thinking about how to best design an online course, programmers who develop online learning products, purchasers who buy educational technology products for their schools, and policy makers who decide which online learning platforms to utilize.

**References**


Allal, L., & Lopez, L. M. (2005). Formative assessment of learning: A review of publications in French. In OECD (Ed.), *Formative assessment: Improving learning in secondary classrooms* (pp. 241-264). France: OECD Publishing.

Allen, Elaine. I., & Seaman, Jeff. (2007). *Online Nation Five Years of Growth in Online Learning*. Needhan, MA: Sloan-C.

Baturo, A., & Nason, R. (1996). Student teachers' subject matter knowledge within the domain of area measurement. *Educational Studies in Mathematics*, 31, 235-268.

Bell, B., & Cowie, B. (2000). The characteristics of formative assessment in science education. *Science Education*, 85, 536-553

Belland, Brian., Glazewski, Krista D., and Richardson, Jennifer C. (2008). A scaffolding framework to support the construction of evidence-based arguments among middle school students. *Education Tech Research Dev*., 56, 401-422.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: putting it into practice*. Buckingham, UK: Open University Press.

Black, P., & Wiliam. D. (1998a). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-74.

Black, P., & Wiliam, D. (1998b). Inside the black box. *Phi Delta Kappan*, 80(2), 139-148.

Bloom, B. S. (1969). *Some theoretical issues relating to education evaluation.* In: TylerRW, (ed.) Educational Evaluation: New Role, New Means. Chicago, IL: University of Chicago Press.

Borasi, R. (1996). *Reconceiving Mathematics Instruction: A Focus on Errors*. Norwood, New Jersey: Ablex Publishing Cooperation.

Buchanan, T. (2000). The efficacy of a world-wide web mediated formative assessment. *Journal of Computer Assisted Learning*, 16(3), 193-200.

Burger, W. F., & Shaughnessy, J. M. (1984) Characterising the van Hiele levels of development in geometry. *Journal for Research in Mathematics Education*, 17, 31-48.

Butler, R. (1987). Task-involving and ego-involving properties of evaluation: effects of different feedback conditions on motivational perceptions, interest and performance. *Journal of Educational Psychology,* 79, 474-482.

Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology*, 58, 1-14.

Chappuis, S., & Stiggins, R. J. (2002). Classroom assessment for learning. *Educational Leadership*, 60(1), 40-43.

Crowley, M. L. (1987). The van Hiel model of the development of geometric thought. *Learning and Teaching Geometry*. 12, 1-16.

Derry, S. J. (1996). Cognitive schema theory in constructivist debate. *Educational Psychologist*. 31, 163-174.

Donovan, M. S., & Bransford, J. D. (2005). Introduction. In M. S. Donovan & J. D. Brandsford (Eds.), *How students learn: History, mathematics, and science in the classroom* (pp.1 28).Washington, DC: National Academic Press.

Dunn, K., & Mulvenon, S. (2009). A critical review of research on formative assessment: The limited scientific evidence of the impact of formative assessment in education. *Practical Assessment, Research & Evaluation*, 14(7), 1-11.

Even, R., & Tirosh, D. (2008). Subject-matter knowledge and knowledge about students as sources of teacher presentations of the subject matter. *Educational Studies in Mathematics*, 29(1), 1-20.

Fuch, L. S., & Fuch, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children*, 53, 199-208.

Geometry Questions. (n.d.). Retrieved from http://www.onlinemathlearning.com/sat-practice-questions-geometry-1.html

Ginsburg, H. (2009). The challenge of formative assessment in mathematics education: Children's minds, teachers' minds. *Human Development* (0018716X), 52(2), 109-128.

Hattie, J., & Timperely, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81-112.

Hechinger, J. (2010, December 7). U.S. teens lag as China soars on international test . Retrieved July 30, 2011, from http://www.bloomberg.com/news/2010-12-07/teens-in-u-s-rank-25th-on-math-test-trail-in-science-reading.html

Heritage, (2010). *Formative assessment and next-generation assessment systems: Are we losing an opportunity?* Washington, DC: Council of Chief State School Officers. Retrieved July 30, 2011 from http://www.ccsso.org/Resources/Publications/Formative_Assessment_and_Next Generation_Assessment_Systems.html

Hogben, L. (1993). *Mathematics for the million: How to master the magic of numbers.* New York, NY: W. W. Norton & Company.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119, 254-284.

Koenig, K. A., Frey, M. C., & Detterman, D. K. (2007). ACT and general cognitive ability. *Intelligence*, doi:10.1016/j.intell.2007.03.005, 1-8.

Li, X., & Li, Y. (2008). Research on students' misconceptions to improve teaching and learning in school mathematics and science. *School Science and Mathematics*, 108(1), 4–7.

Lipnevich, A., & Smith, J. (2009). Effects of differential feedback on students' examination performance. *Journal of Experimental Psychology*, 15(4), 319-333.

Mamba, F. T. (2011). An investigation into students' misconceptions in linear equations in public secondary schools of malawi: The case of the south eastern education division. Hiroshima, Japan: Hiroshima University.

Mayer, R. E.(1984). *A cognitive analysis of mathematical problem-solving ability*. New York, NY: R.J. Sternberg (ed.).

Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends In Cognitive Sciences*, 5(8), 349-357.

Office of Economic Co-operation and Development. (2005). *Formative assessment: Improving learning in secondary classrooms*. Paris, FR: OECD Publishing.

Piaget, J. (1952). *The origins of intelligence in children*. New York: International Universities Press.

Pind, J., Gunnarsdottir, E., & Johannesson, H. (2003). Raven's standard progressive matrices: New school age norms and a study of the test's validity. *Personality and Individual Differences*, 34, 375–386.

PISA. (2011). Retrieved July 30, 2011, from http://www.pisa.oecd.org/pages/0,2987,en_32252351_32235731_1_1_1_1_1,00.html

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, 28, 4-13

Reveles, J. M., Kelly, G. J., & Durán, R. P. (2007). A sociocultural perspective on mediated activity in third grade science. *Cultural Studies of Science Education*, 1, 467-495.

Sadler, D. (1989). Formative assessment and the design of instructional systems. *Instructional*

*Science*, 18, 119-144.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, and M. Scriven (Eds.), *Perspectives of curriculum evaluation*, Volume I (pp. 39-83). Chicago, IL: Rand McNally.

Shaughnessy, J. M. (1985). Problem-solving derailers: The influence of misconceptions on problem-solving performance. In E.A. Silver (Ed.), *Teaching and learning mathematical problem solving* (pp.399-415). Hillsdale, NJ: Erlbaum.

Smith, J., Disessa, A., & Roschelle, J. (1993). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences*, 3(2), 115-163.

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*. Retrieved July 5, 2011 from http://www.pdkintl.org/kappan/k0206sti.htm

Tanner, H. & Jones, S. (2003). Self-efficacy in mathematics and students' use of self-regulated learning strategies during assessment events. Proceedings from 27th International Group for the Psychology of Mathematics Education. PME: HI.

Taras, M. (2009). Summative assessment: the missing link for formative assessment. *Journal of Further and Higher Education*, 33(1), 57-69.

Tharp, R. G., & Gallimore, R. (1988). *Rousing minds to life: Teaching, learning, and schooling in social context*. Cambridge, MA: Cambridge University Press.

Usiskin, Z. (1982). *Van Hiele levels and achievement in secondary school geometry*. Chicago, IL: Chicago Press.

van Hiele, P. M. (1999). Developing geometric thinking through activities that begin with play. *Teaching Children Mathematics,* 6, 310-316.

Vygotsky, L. S. (1978). *Mind and society: The development of higher mental processes.* Cambridge, MA: Harvard University Press.

Wang, K., Wang, T., Wang, W., & Huang, S. (2006). Learning styles and formative assessment strategy: Enhancing student achievement in web-based learning. *Journal of Computer Assisted Learning*, 22(3), 207-217.

Waterman, S. S. (2010). *Assessing middle and high school mathematics and science: Differentiating formative assessment*. Larchmont, NY: Eye On Education.

Whiting, B., Van Burgh, J. W., & Render, G F. (1995). *Master learning in the classroom*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Who Are We. (2011). Retrieved July 30, 2011, from http://www.dcccd.edu/About%20DCCCD/Who%20We%20Are/Pages/default.aspx

Wiliam, D. (1999). Formative assessment in mathematics part 2: Feedback. *Equals: Mathematics and Special Educational Needs* 5(3) 8-11.

Wiliam, D., & Leahy, S. (2007). A theoretical foundation for formative assessment. In J. McMillan (Ed.), *Formative classroom assessment: Theory into practice*. NewYork, NY: Teachers College Press.

Wininger, R. S. (2005). Using your test to teach: Formative summative assessment. *Teaching Psychology*, 32(2), 164-166.

Wolters, C. A., & Rosenthal, H. (2000). The relation between students' motivational beliefs and their use of motivational regulation strategies'. *International Journal of Educational Research*, 33, 801-820.

**Appendix**

Student Interest Questionnaire

| | Strongly Agree | Agree | Not Sure | Disagree | Strongly Disagree |
|---|---|---|---|---|---|
| I found the material we covered interesting. | ○ | ○ | ○ | ○ | ○ |
| I would like to participate in an experiment like this again. | ○ | ○ | ○ | ○ | ○ |
| I would recommend this experience to others. | ○ | ○ | ○ | ○ | ○ |
| This was very stimulating. | ○ | ○ | ○ | ○ | ○ |
| I am more interested in mathematics as a result of this experience. | ○ | ○ | ○ | ○ | ○ |
| I enjoyed the experience. | ○ | ○ | ○ | ○ | ○ |
| I would like to take an online class like this the future. | ○ | ○ | ○ | ○ | ○ |
| I enjoyed this math experience more than other math experiences in the past. | ○ | ○ | ○ | ○ | ○ |
| I have a new-found interest in mathematics. | ○ | ○ | ○ | ○ | ○ |
| I feel this type of experiment was successful. | ○ | ○ | ○ | ○ | ○ |