Teachers’ Understanding and Use of Formative Assessments in the Elementary Mathematics Classroom

The Harvard community has made this article openly available. **Please share** how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th>Citation</th>
<th>Harris, Steven E. 2016. Teachers’ Understanding and Use of Formative Assessments in the Elementary Mathematics Classroom. Doctoral dissertation, Harvard Graduate School of Education.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:27112699">http://nrs.harvard.edu/urn-3:HUL.InstRepos:27112699</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Teachers’ Understanding and Use of Formative Assessments
In the Elementary Mathematics Classroom

Steven E. Harris

Thesis Committee
Eileen McGowan
James Honan
Stuart Berger

A Thesis Presented to the Faculty
Of the Graduate School of Education of Harvard University
In Partial Fulfillment of the Requirements
for the Degree of Doctor of Education

2016
DEDICATION

To my brother, Jeffrey Edward Harris,

who always encouraged me to follow my dreams.
# TABLE OF CONTENTS

## ABSTRACT

vi

## CHAPTER 1: INTRODUCTION AND REVIEW OF THE LITERATURE

Review of Literature 4
- History of Formative Assessment 4
- Current State of Formative Assessment 8
- Creating a Definition 12

Chapter Summary 12

## CHAPTER 2: RESEARCH DESIGN

Research Questions 14

Site and Participant Selection 15
- Table 1: Central School District Teachers 15

Data Collection 16
- Interviews 16
- Classroom Observations 16

Data Analysis 17

Validity 20

## CHAPTER 3: TEACHER PERCEPTIONS OF FORMATIVE ASSESSMENT

Teachers’ Definitions of Formative Assessment 21
- What Happens After the Initial Formative Assessment Task 24
- Table 2: Teachers’ Definitions and Examples by Theme Level 25

Types of Formative Assessments Teachers Reported Using 26
- Table 3: Reported Types of Formative Assessments 26
- Table 4: Reported Types and Whether Imbedded in Lesson Plans 29

Teachers’ Self Rating of Their Use of Formative Assessment 29
- Table 5: Teachers’ Self Ratings, Steps to Improve and Barriers 31

Chapter Summary 32
ABSTRACT

In 1998 Paul Black and Dylan Wiliam published the article, Inside the Black Box: Raising Standards Through Classroom Assessment (Black & Wiliam, 1998b). They asserted that formative assessments were the strongest way of raising student achievement. There are a number of empirical studies that document positive impacts of formative assessment on student learning (Brookhart, 2004; Allal & Lopez, 2005; Köller, 2005; Brookhart, 2007; Wiliam, 2007; Hattie & Timperley, 2007). There are also critics of much of the existing research (Shute, 2008; Dunn and Mulvenon, 2009; Bennett, 2011; Coffey Hammer Levin and Grant, 2011).

The literature points to the need for more research in this area. Shavelson (2008), who looked at formative assessment in the science classroom stated, “[a]fter five years of work, our euphoria devolved into a reality that formative assessment, like so many other education reforms, has a long way to go before it can be wielded masterfully by a majority of teachers to positive ends. This is not to discourage the formative assessment practice and research agenda.”

In this study I examined how teachers understand formative assessment in relation to their instruction, and how they actually implement formative assessment in their math classrooms. I used a thematic analysis research design, analyzing interviews, and observation recordings and field notes. I created a Depth of Implementation Framework, based on both a definition of formative assessment constructed from the review of literature and on the data gathered from teachers, to help make sense of the interplay between teachers’ understanding and use of formative assessment. Based on the data, teachers’ use of formative assessment was characterized as deep, developing or
superficial.

Teachers’ understanding of formative assessment, especially the definitions that they constructed for themselves, had an impact on both how they used formative assessment and how they saw themselves improving their use of formative assessment.
CHAPTER 1: INTRODUCTION AND REVIEW OF THE LITERATURE

In 1998 Paul Black and Dylan Wiliam published the research review, *Inside the Black Box: Raising Standards Through Classroom Assessment* (Black & Wiliam, 1998b) where they review 250 articles and book chapters. The last paragraph of their review read:

> There is a body of firm evidence that formative assessment is an essential component of classroom work and that its development can raise standards of achievement. We know of no other way of raising standards for which such a strong prima facie case can be made. (p. 147)

This bold assertion became ground zero for the belief that educators can use formative assessment as a tool to raise student achievement.

Over the eighteen years since its publication, Black and Wiliam’s (1998b) article has been cited countless times as educators have studied, researched and ruminated about formative assessments (Stiggins, 2005; Aschbacher and Alonso, 2006; Ruiz-Primo and Furtak, 2006; Popham, 2008; McManus, 2008; Harlen & Crick, 2003; Heritage, 2007). The use of formative assessments in school districts is widespread, with over $600 million being spent on resources and professional development in 2007 (Cech, 2008).

There are a number of empirical studies that document positive impacts of formative assessment on student learning (Black & Wiliam, 1998; Nyquist, 2003; Brookhart, 2004; Allal & Lopez, 2005; Köller, 2005; Brookhart, 2007; Wiliam, 2007; Hattie & Timperley, 2007). There are also critics of much of the existing research (Shute, 2008; Dunn and Mulvenon, 2009; Bennett, 2011; Coffey Hammer Levin and Grant, 2011). One of the main criticisms is the lack of definition.

The vagueness of the constitutive and operational definitions directly contributes to the weaknesses found in the related research and dearth of empirical evidence identifying best practices related to formative
assessment. Without a clear understanding of what is being studied, empirical evidence supporting formative evidence will more than likely remain in short supply (Dunn and Mulvenon, 2009, page 2).

The critical literature also takes issue with much of the methodology around measurement, number of variables, and multiple effects from the same study (Shute, 2008; Dunn and Mulvenon, 2009; Bennett, 2011; Coffey Hammer Levin and Grant, 2011; Kingston and Nash, 2011).

Shavelson (2008), who looked at formative assessment in the science classroom stated, “[a]fter five years of work, our euphoria devolved into a reality that formative assessment, like so many other education reforms, has a long way to go before it can be wielded masterfully by a majority of teachers to positive ends. This is not to discourage the formative assessment practice and research agenda.” p. 294. The literature points to the need for more research in this area, and specifically, more research in classrooms (Dunn and Mulvenon, 2009). “Very few studies exist on how interim assessments are actually used, by individual teachers in classrooms, by principals, and by districts.” (Goertz, Oláh, and Riggan, 2009).

The current conditions in education policy in the United States also point to the need for more research around instructional strategies that are effective in raising student achievement. With the demise of No Child Left Behind, and the advent of the Every Student Succeeds Act, which was signed into law in December 2015 and will take effect in the 2017-18 school year, the nation is going through a very large educational policy shift. The federal government is stepping back from education policy and enhancing the states’ authority over education policy. This means education policy makers in every state will be searching for effective methods and models to implement in their states’
schools that carry the promise of improving outcomes for their students. The research of educational practice, including formative assessment research, will take on renewed importance in the next few years.

For this research I am looking specifically at how teachers understand and use formative assessments. In the Venn diagram of instructional research, formative assessment overlaps with, or is tangential to, several other instructional methodologies and research areas. I raise this issue to assist the reader in understanding the specific focus of this research paper.

Formative assessment is related to teaching for understanding. Because of the absence of a universal definition of formative assessment, it is difficult to articulate the exact difference between teaching for understanding and formative assessment, but generally speaking, the use of formative assessments is a subset of teaching for understanding which encompasses strategies outside of formative assessment such as providing clear standards (Darling-Hammond, Barron, et al., 2015).

This research looks at formative assessment from the perspective of how teachers understand and use it. This is closely related to the research around teacher sense-making. The distinction between the two is that the teacher sense-making research focuses on reforms; times when teachers are asked to implement something new to them (Luttenberg, Veen, & Imants 2013; März & Kelchtermans. 2013). Formative assessment is not a new reform, but rather an ill-defined methodology that teachers are already using, and how teachers use and understand formative assessment is the focus here.

In this first chapter I review the literature on formative assessment. In chapter 2 I review the research design. In chapter 3 I report on teachers’ perceptions around their
use of formative assessments, and in chapter 4 I review teachers’ actual implementation
of formative assessment in their classrooms. In chapter 5 I examine teacher’s use of
formative assessment through the Depth of Implementation framework. In chapter 6 I
review factors impacting teachers’ use of formative assessments and in the final chapter,
7, I review the limitation and implications of this research.

Review of the Literature

History of Formative Assessment

The concept of formative assessment is not a new one to educators. While I
suspect that good teachers have always taken measures in their classrooms to determine if
and to what degree students understand the materials, it was not until the mid 1960’s that
the concept and phrase began appearing in educational literature. Lindvall, in his book
Measuring Pupil Achievement and Aptitude (1967), expresses the idea that it is the
teacher’s role to give assessment throughout the course of instruction to improve the
processes of teaching and learning. Lindvall believed that this could be done formally
through written exams, and informally through questioning, observation and homework.
Lindvall, however, did not use the phrase “formative assessment.”

Michael Scriven first used the word “formative” as part of an assessment process
in a 1967 article. Scriven was specifically speaking of the development and evaluation of
school curriculum, saying “formative evaluation involves the collection of appropriate
evidence during the construction and trying out of a new curriculum in such a way that
revision of the curriculum can be based on this evidence” (p. 63).

The concept of formative assessment gained wider recognition in the education
community with the publication of Benjamin Bloom’s book, Handbook for Formative
and Summative Evaluation of Student Learning (Bloom, Hastings, & Madaus, 1971). Bloom, et al (1971) are more explicit in their description of formative evaluation stating, “Formative evaluation is for us the use of systematic evaluation in the process of curriculum construction, teaching, and learning for the purpose of improving any of these three processes (p. 118).

Bloom, et al (1971) took Scriven’s definition of formative evaluation and more clearly define it for teaching and learning. They contrasted their description with summative evaluation and pointed out the timing differences that Scriven started with. Bloom et al (1971) go on to say that with formative evaluations there should be no judgment as there is with summative evaluations. Formative evaluations are for improving curriculum development, teaching and learning.

Bloom et al (1971) also make explicit that formative evaluations are to be used by teachers and students. It is not just the adults in the education setting who should use the formative evaluation to improve teaching and learning. Students should also be using formative evaluations to improve learning.

Over the next decade, a number of studies were published which tended to support the notion that teachers’ use of formative assessments in the classroom improved student achievement on standardized tests (Fiel & Okey, 1974; Green, 1975; Block & Burns, 1976). Studies by Mirkin & Deno (1979, 1980) supported the effectiveness of formative assessments with special education populations.

In the 1980s the research began to refine the formative assessment process finding that feedback to students is an essential part of formative assessments (Berliner, 1980; Natrielo, 1987; Crooks, 1988). Sadler (1989) took this one step further, finding that
teachers should give specific feedback, which includes what students are doing well, things students might do to improve, and give students a sense of where they should be in the understanding of the content.

Airasian (1994) and McMillan (1997) continued the research in the 90s, both focusing on the interactive nature of formative assessments (between teacher and student) and emphasizing the fact that formative assessments should not be used to assign grades to student work (Guskey (1996))

In 1998, Black and Wiliam published a pair of articles (1998a & b) that have become the seminal source of evidence of the effectiveness of formative assessments in raising student achievement. The first piece, Assessment and Classroom Learning (1998a), is a literature review that looks at some 250 articles, including about 20 quantitative studies. Black and Wiliam presented eight studies in their article, and in their analysis of the studies they conclude:

For public policy towards schools, the case to be made here is firstly that significant learning gains lie within our grasp. The research reported here shows conclusively that formative assessment does improve learning. The gains in achievement appear to be quite considerable, and as noted earlier, amongst the largest ever reported for educational interventions. (1998a, p.61).

They cite achievement gains between .4 and .7 standard deviations (p. 61).

During the time that the research on formative assessments was leading to the Black and Wiliam peak, assessments were becoming more present on the policy side of education in the form of summative assessments. During the 1970s and 80s in the United States, states were beginning to implement statewide testing programs, and in 1983 A Nation at Risk was published by the US Department of Education (Gardner, 1983). The findings of this
report were critical of the education system in the United States and suggested a number of reforms to improve schools, which moved the nation to seriously begin exploring education reform. The reforms tended to have an assessment component and schools moved from using assessments as a way to gather information about students to a way of ranking students by achievement (Eisner, 2002). Along with standardized testing by individual states, the National Assessment for Educational Progress board was created in 1988, and NAEP testing began in 1992.

In 2002 the No Child Left Behind Act was enacted, and states were now compelled by law to assess every school and rank them based on their standardized test scores. Because of this mandate, some researchers believe that many educators see assessments as compliance with policy instead of as an instrument for academic improvement (Stiggins, 2004; Gardner, 2011). This also put educational focus on the high stakes summative assessments instead of the formative assessments.

Aschbacher and Alonso (2006) expanded on the idea that formative assessments do not have to be formal. Specifically, they examined teachers’ use of elementary science notebooks to understand student thinking, and found that this practice produced gains in student achievement. Ruiz-Primo and Furtak (2006) also looked at formative assessments through questioning and games. They found that these formative assessment methods allowed teachers to adjust their instruction and provide helpful student feedback.

In his book Transformative Assessment, Popham (2008) begins with Black and Wiliam’s view that formative assessments raise student achievement and attempts to operationalize formative assessments for use by educators. He not only advocates formative assessments as a teaching tool, he broadens its use as a way to change the
culture of a school through teacher learning communities centered on the sharing and professional development of formative assessment teaching strategies. Popham (2008) also puts more emphasis on the relationship between the teacher and the student in the formative assessment process.

**Current State of Formative Assessment**

The use of formative assessment as a tool to raise student achievement has received a fair amount of literary discussion since Black and Wiliam’s claims that it is the strongest method of raising standards of achievement (1998b). Formative assessment has not, however, born out the widespread achievement gains that Black and Wiliam (1998b) envisioned possible.

According to The Nation’s Report Cart (National Center for Education Statistics, 2011a), which reports the results of the National Assessment of Educational Progress (NAEP) exam, in the U. S., 4th grade reading scores have risen .8% and 8th grade reading scores have risen 1% since 1992. In Mathematics, 4th and 8th grade scores have risen 5.6% and 4.2% respectively since 1992 (National Center for Educational Statistics, 2011b). It is clear by these numbers that the nation has not yet found or implemented the “fix” to raising academic achievement.

Not only has formative assessment not shown the promising results reported by Black and Wiliam, but in recent years Black and Wiliam’s (1998a) study of formative assessment has come under critical review (Dunn, 2009; Bennett, 2011; Coffey Hammer Levin & Grant 2011).

Dunn’s (2009) criticism of Black and Wiliam’s (1989a) study is based largely on research methodology. In the eight studies that Black and Wiliam (1998a) rely on there
are issues of technically inadequate dependent measures, uncontrolled examiner expectancy, unchecked fidelity of treatment, inappropriate statistical unit of analysis, small sample size, and inappropriate control groups. In summarizing Black and Wiliam’s work, Dunn (2009) states: Collectively, the eight articles that Black and Wiliam (1998a) presented to *conclusively* show that formative assessment significantly improves student achievement do not support such a conclusion. A more appropriate conclusion may have been that more research needed to be done. (p. 7)

Dunn (2009) goes on to say:

Stiggins (2005) stated that “to use assessment productively to help achieve maximum student success, certain conditions need to be satisfied” (p. 4). The authors agree with this statement, but also posit two conditions that were not stated by Stiggins. First, a clear and shared lexicon must be established and shared among all educational stakeholders to lead to more productive communication among teachers, researchers, policy makers, parents, and students. Finally and most importantly, a sound research-validated framework for best practices in formative assessment and formative evaluation must be established to ensure maximum benefits for all those involved.

In the article *Formative Assessment: A Critical Review*, Bennett (2011) identifies six issues that he believes have been absent from the dialogue around formative assessments. Bennett (2011) identifies those issues as “the definitional issue, the effectiveness issue, the domain dependency issue, the measurement issue, the professional development issue, and the system issue” (p. 5).

In describing the definitional issue Bennett states:

Definition is important because if we can’t clearly define an innovation, we can’t meaningfully document its effectiveness. Part of that documentation needs to be an evaluation of whether the formative assessment was implemented as intended, which we cannot accomplish if we don’t know what was supposed to be implemented. Similarly, if we can’t clearly define an innovation, we can’t meaningfully summarize results across studies because we won’t know which instances to include
in our summary. Last, we won’t be able to transport it to our own context, for how will we know the characteristics on which to focus in doing the transport? (p. 8)

Bennett (2011) describes the effectiveness issue by pointing out many of the same methodology issues in Black and Wiliam’s work as Dunn (2009) does. The domain dependency issue speaks to whether the generalized discussion of formative assessment across domains is maximally effective, when deep domain understanding is believed to be a necessary component of helping students learn. Each domain has different processes and strategies that should be taken into account in the formative assessment process. (Bennett, 2011)

Bennett’s (2011) measurement issue questions the strength of teacher’s ability to interpret formative assessment evidence. When educators see evidence of an error in the student’s thinking, they generally have to infer why that error occurred. Formative inferences are not only subject to uncertainty, they are also subject to systematic, irrelevant influences that may be associated with gender, race, ethnicity, disability, English language proficiency, or other student characteristics. Put simply, a teacher’s formative actions may be unintentionally biased. (Bennett, 2011, pp.17-18).

The professional development issue raises the need to include measurement fundamentals and deep domain understanding in effective formative assessment implementation. Bennett’s (2011) final issue is the system issue. If the system of education where formative assessments are being used is not supportive and aligned with the initiative, then its effectiveness will be limited.

Despite Bennett’s criticism, formative assessment research continues to find positive effects on student achievement, and the research continues to support its use in
classrooms. In a 2015 meta-analysis of 40 studies, researchers found positive effects when feedback to students contained explanations instead of just whether a student was right or wrong (Van der Kleij, Feskens, & Eggen, 2015). Conley, (2014) in a report commissioned by Jobs for the Future states that, “the pursuit of deeper learning will require much greater emphasis on formative assessments.” (p. 22). There is also a growing trend to replace the phrase formative assessment with the phrase, “assessment for learning,” which is more descriptive than formative assessment (Stiggins, 2014). Although the terminology and definitions are still in flux, the general concept of formative assessment, that a teacher should understand and respond to student thinking about a subject, is the phenomenon that still carries hope.

While Black and Wiliam’s initial optimism for formative assessment’s potential has been tempered a bit with a more critical eye on research methods, there is still optimism for formative assessment and its use in the classroom. The research below will help add to educators’ knowledge about actual use of formative assessments in the classroom and give some insight into how to improve that use.

**Creating a Definition**

The variety of definitions raises several questions. Is formative assessment the instrument, the process, or both? Is a single definition of formative assessment sufficient to describe what is happening? Are these definitions as helpful as they can be for the practitioners who are trying to use them?

Addressing these questions, I begin with instrument vs. process. The problem is not whether formative assessment is one or the other. The problem is that historically the literature has used the same term to mean one, or the other, or both. At this point it would
be entirely too cumbersome to create another word for either the instrument or the process. Instead, I will not answer the question directly, but will use formative assessment as both instrument and process, and will be more specific within the term formative assessment. For example, I will refer to the instrument or task used to gather evidence as the formative assessment instrument, task or questions instead of as the formative assessment itself.

I also agree with Shavelson et al (2008) and believe that the definition should be broken up. The three distinctions that they make certainly make sense in the context of a classroom, and I will use this as part of my own definition and framework.

I believe the definitions should contain the elements that are generally necessary to make a formative assessment a formative assessment. The research repeatedly mentions gathering information/evidence from the student, using this information to understand the student’s thinking, making decisions regarding the modification of instruction, giving students feedback, and engaging them in their own learning. If these are in fact necessary elements, then they should be reflected in the definition.

My working definition is as follows:

Formative Assessment is made up of the tools and processes used during teaching and learning to raise student achievement by gathering information to understand student thinking, providing helpful student feedback, modifying our instruction based on that information, and engaging students to be active in their own academic success. Information may be gathered through a formal written instrument or informally through planned activities and discussion or through unplanned opportunities such as discussion between and amongst students.

Chapter Summary

Although formative assessment as a concept has been bandied about in education circles since the mid-sixties, it is clear through the research that there has not been
enough rigorous research and development of best practices to create a finely tuned process that is replicable in classrooms across America. Black and Wiliam (1998b) gave a strong endorsement of formative assessment’s effectiveness, which went unquestioned until recently. As the latest research begins to question Black and Wiliam’s (1998b) assumptions, a window of opportunity to start anew and develop best practices with formative assessments has emerged.

In synthesizing the research, I believe that there is enough information to create a workable definition, which I have done, and I use this definition as the starting point for creating a Depth of Implementation framework, which will be described in greater detail in Chapter 5.
CHAPTER 2: RESEARCH DESIGN

In this study I looked at teachers’ understanding and use of formative assessment. Specifically, I sought to understand how teachers view and understand their own use of formative assessment in the classroom. Formative assessment is not a new methodology in education, and it is a generally accepted practice in classroom instruction, however, the professional development around formative assessment tends to be general, working from the assumption that teachers know what it is.

The district in this study did not train teachers in a specific methodology of formative assessment. The district generally supported the use of formative assessment but operated under the assumption that teachers knew what it was and how to use it. Teachers were free to use their own understanding of formative assessment in their classrooms. If teachers had specific questions or needs around the use of formative assessment the administrators in the district were willing to provide individual and specific supports.

Research Questions

The following questions guided my study:

1. How do elementary math teachers describe formative assessment?
2. How do elementary math teachers use formative assessment in relationship to their teaching math?
3. What obstacles, if any, do elementary math teachers perceive to using formative assessment in teaching math?
Site and Participant Selection

The school district used for this study is Central School District.\(^1\) Central School District is a small suburban district in the Midwest. The district has 5 schools and a student population of 2,100. The student body is 93% white, 3% Latino, and .5% black and Asian, and 24.4% of students qualify for free and reduced lunch. The district’s student outcomes are closely aligned to the state averages on the state’s standardized tests. The district is essentially in the middle of the pack in terms of student performance for school districts in the state, and represents an “average” district. I used “purposeful selection” (Maxwell, 2013, p. 97) to choose this district because I wanted a district whose characteristics were generally average, so this study might reflect what is happening in a large cross section of school districts in the country.

Individual teachers were identified for the study through the Assistant Superintendent for Academics, who sent my invitation letter (see Appendix A: Invite Letter) through email to all of the 3\(^{rd}\) through 6\(^{th}\) grade teachers along with a personal note asking for volunteers. Thirteen of the initial sixteen teachers who were invited volunteered to be a part of the study. The study participants are listed below in Table 1.

Table 1

<table>
<thead>
<tr>
<th>Name</th>
<th>School</th>
<th>Grade Taught</th>
<th>Years Teaching</th>
</tr>
</thead>
<tbody>
<tr>
<td>Julie Jones</td>
<td>Central Elementary</td>
<td>3rd</td>
<td>19</td>
</tr>
<tr>
<td>Natalie Night</td>
<td>Central Elementary</td>
<td>3rd</td>
<td>15</td>
</tr>
<tr>
<td>McKenzie Orr</td>
<td>Central Elementary</td>
<td>3rd</td>
<td>5</td>
</tr>
<tr>
<td>Trish Tisby</td>
<td>Central Elementary</td>
<td>3rd</td>
<td>2</td>
</tr>
</tbody>
</table>

\(^1\) All district, school, teacher and student names are pseudonyms.
Data Collection

I collected data through interviews and through observations of instruction.

Interviews

I began the first interview with each participant by walking through the consent letter (See Appendix A Consent Letter). After I obtained their signature and gave them a copy, I began asking the questions that I had prepared before the interview (See Appendix B Interview Protocol). I audio recorded the first interviews, with the teachers’ permission, and began by briefly asking teachers about their biographical background to get them comfortable with the interview process. I then delved into their understanding of formative assessments and how they use them in their math class. Each of the interviews lasted between 40 minutes and an hour, and following each of these interviews a transcription of the interview was created.

The IRB Approval Letter for this study can be found in Appendix C.

Classroom Observations

In the second step of my data collection, I observed each teacher teach two 80-
minute lessons. I conducted the observations in two rounds. After observing every teacher in the first round, I observed every teacher again, so no two observations of a teacher were consecutive lessons. In each of the observations I took detailed field notes to capture what is happening in the classroom during the observation, using an open observation form (See Appendix D). As I observed, I documented the time, and anything that might be a formative assessment. I documented what the teacher was doing and what the students were doing regard what happens around that assessment. If the formative assessment involved dialogue, I wrote it down, and I also audio recorded each of the classroom observations and transcribed the recordings, matching up the time stamp in the recordings to my documented times to make sure my notes and recordings were in sync.

While observing, I was only trying to capture what was actually happening. At that point, I was not concerned about whether teachers’ actions matched their descriptions from their interviews, or whether the formative assessment was effective. I was simply documenting the formative assessment.

On three occasions I asked teachers follow up questions based on what I had observed in the classroom. In each of these cases, the questions were asked immediately after the classroom observation and in each case the questions were to clarify something that happened in class. I audio recorded the questions and answers and took field notes.

**Data Analysis**

I used thematic analysis (Braun and Clarke, 2006) to analyze the data. Braun and Clarke outline six phases of data analysis which are data immersion, code generation, finding overarching themes, seeing patterns in the theme, creating story about the data
from the themes and reporting. Most of the literature on formative assessment looks at the application of a theoretical model. I was not trying to fit the data into a framework or hypothesis, but rather, I sought to understand what was happening with regard to formative assessments from the teachers’ point of view, so thematic analysis was an appropriate method.

As I collected the data, I transcribed the recordings of my interviews. I also transcribed the audio recordings of the classroom observations and combined these transcriptions with my field notes into one document. I repeatedly reviewed the transcripts to become more familiar with what was contained in the data.

While I had created a definition of formative assessment from the research, I did not use it in the initial coding of the data. I used emic coding to capture the teachers’ understandings of formative assessment (Charmaz, 2009). I started the first round of coding by identifying in-vivo codes for how teachers defined and understood their own use of formative assessment (Saldana, 2015). Because there is not consistent language around formative assessment, I assumed that teachers would use a variety of words choices to describe their own understanding of formative assessment. From these initial in-vivo codes, I began looking for themes across teachers. I looked for themes in how teachers described their use of formative assessment, in how they used tasks and questions to initiate formative assessments and what processed they employed, if any, as follow-up to the formative assessment tasks or questions.

As I pulled themes from the data, I kept my first two research questions in mind. I thought about each question individually, and how the themes helped to explain the understanding and the use of formative assessments and I thought about the questions
jointly- whether the themes from the descriptive understanding and the themes from the actual use matched up and were consistent.

Using the data-generated themes I created a depth of implementation framework to give a meaningful descriptive representation of the pertinent data regarding teachers’ use and understanding of formative assessment. This allowed me to compare teachers across each, or several themes and to combine the themes to look at the use of formative assessment holistically. I repeatedly reviewed the data around each of the themes for each teacher to determine the range of elements within a theme and after reviewing the teachers’ elements across all of the themes I found that the data describing the teachers’ use of formative assessment fell into three levels of depth which I identified as superficial, developing, and deep implementation.

I initially leveled the data based on the definition created through the review of literature, but after reviewing the data repeatedly there was a clear difference between the levels of implementation based on my definition of formative assessment and the teachers’ self perceptions of their levels of implementation. To further analyze this difference I also leveled each teacher’s data based on their own verbal definition of formative assessment.

After the depth of implementation analysis was done, I went back into the data to identify the barriers to implementation in response to my third research question. I took teachers at their word when they verbally identified barriers, but I also looked for patterns in the data between teachers’ verbally identified barriers and their actual use of formative assessment to identify other barriers.
Validity

As the researcher I am analyzing the data, and that analysis necessarily goes through my own filters (Creswell, 2007). While I was trying to understand the phenomenon of the teacher’s use of formative assessment, there is a threat of bias in the fact that I was starting with a definition, which may alter how I see the data.

I also come to this research with the experience of having practiced law for a number of years as a trial attorney. Taking a deposition is qualitative research. I always entered depositions with a theory of my case, but I also understood that the deposition was not about filling in my theory, but rather understanding the deponent’s story as completely as possible. The effect of not getting the full story was that you might be surprised in trial with information that you didn’t ask about, which was never a good thing. Having this life experience helps me to understand and minimize my own biases when seeking to understand someone else’s story.

I collected data from three or four different points, and in two different methods for each of the participants (first interview, two observations, and second interview when needed). I used triangulation with these data points, along with my own field notes which I took at all of the interviews and observations, to help minimize my own bias.

I also shared my work for peer review with a doctoral candidate colleague. We have known each other for six years, have worked together for the last year, and have read and discussed each others doctoral work for the last year. She is very familiar with me, and my work, and will assist in minimizing my biases in this research.
CHAPTER 3: TEACHERS’ PERCEPTIONS OF FORMATIVE ASSESSMENT

“It is important for research on educational reform to attend closely to the actual implementation process and to how it is perceived and motivated by the people who enact it in order to understand and explain the reform outcomes.” (März & Kelchtermans. 2013). In this chapter I examine teachers’ perceptions of their implementation of formative assessments in their math classrooms. Specifically I look at data regarding teachers’ definitions of formative assessment, the types of formative assessments that teachers use, and the teachers’ self-ratings of their own effectiveness using formative assessments in the classroom.

Teachers’ Definitions of Formative Assessment

I asked all of the teachers what types of professional development and training that they had received using formative assessments. Each of the teachers indicated that somewhere in their background, either in college or through professional development in a school district that they had touched on assessments in a general way. None of the teachers had received specific or ongoing professional development or training on the use and implementation of formative assessments. One of the teachers expressed her use of formative assessment as “developing naturally” from teaching and curriculum writing. (Natalie Night Interview, August 14, 2014). Another teacher described the training that she had received by saying, “the professional development says to do it [formative assessment] but it did not say a lot about how to do it.” (Trish Tisby Interview, August 13, 2014).

It was clear from this line of questions that none of the teachers had adopted someone else’s specific definition of formative assessment from a training program or
professional development. The district also did not have a specific definition that it asked teachers to adopt. Julie Jones, who has been with the district 19 years said, “We talk about it generally in professional development meetings with other things, but we don’t have professional development specifically about formative assessment.” (Julie Jones Interview, August 15, 2014).

All thirteen of the teachers expressed an understanding of formative assessment and also expressed the belief that they were using it in their classrooms. They all also expressed a general belief that formative assessment was a process that allowed them to better assist students in learning math. When I asked how they define formative assessment, I found that teachers gave answers that fell into three themes. The first theme was simply, “do students get it?” The second theme was, “do student get it, and do I need to slow down or reteach?” The third theme was, “do students get it, and do I need to modify my instruction.

In the first theme, teachers gave definitions of formative assessment that were just about the student. Ten of the teachers defined formative assessment as a process that only involved determining whether a student understood a concept or not. For example, Cecilia Reyes defined it as, “[c]hecking to see if students know what they are doing.” (Cecilia Reyes Interview, August 21, 2014). Trish Tisby responded by saying, “seeing if a student can use the full vocabulary.” (Trish Tisby Interview, August 13, 2014). Bill Hollister defined formative assessment as, “checking to see where the kids are in the process of learning.” Micah Norris’ definition was a simple, “seeing if students get it.” (Micha Norris Interview, August 22, 2014).
Two of the teachers gave definitions along the second theme that involved both student understanding and a component of teacher action. Those two teachers referenced either pace or re-teaching. Natalie Night defined formative assessment as, “the way I gauge student understanding, which drives the pace of my instruction.” (Natalie Night Interview, August 14, 2014). Pat Paulson defined it as, “constantly walking and seeing if kids are getting it, and re-teaching if necessary.” (Pat Paulson Interview, August 14, 2014).

There was only one teacher whose definition defined the third theme. Julian Keller defined formative assessment as, “a quick way of making sure every student understands what you’re teaching and allows you to modify your instruction as you go.” (Julian Keller Interview, August 13, 2014). The definitions in both the second and third themes contained some action by the teacher. The distinction between these two themes is whether that action is an actual change to the teachers’ instruction based on the information received from the students.

In the second theme the teachers indicated that they would change pace or reteach the material. Changing the pace of instruction does not mean that the teacher would do anything differently in her instruction except slow down or speed up. The actual instruction does not necessarily change. The same is true of re-teaching. Without anything more, re-teaching implies that the same instruction happens again. Julian Keller specifically used the words “modify instruction.” This is a clear indication that the instruction has to change, and that clear indication (as opposed to an implied indication) is what distinguished the second from the third themes.
What Happens After the Initial Formative Assessment Task

To gain more insight into how teachers defined formative assessment, I asked teachers to give an example of what they would do “if you have students in your class who indicate that they didn’t understand something during class?” The teachers’ answers produced three themes, each involving some type of review with students. The first theme was to simply reteach the material. The second theme was to reteach the material using different instruction, and the third theme was to modify instruction based on an understanding of student thinking.

Nine of the teachers were in the first theme and described a simple process of re-teaching, such as, “I would pull them out for one on one or small group work,” (Natalie Night Interviews, August 14, 2014) or “I would walk them through the process again” (Pat Paulson Interviews, August 14, 2014). These nine teachers did not indicate that the instruction would look any different in the re-teaching than it did the first time for the whole group instruction.

Two of the teachers gave responses that explicitly included some type of modification to instruction in the re-teaching, which defined the second theme. Those two teachers said, “I would go back to the kids later and try to rephrase, try to explain it in a different way,” (Julian Keller, August 13, 2014), and, “I would review, or physically move pieces, or reteach in a totally different way with pictures.” (Trish Tisby Interview, August 13, 2014).

The third theme came from two of the teachers’ responses, where they explicitly started that in addition to modifying instruction they would try to figure out student
thinking. Howard Stenzel explained, “I would have the students walk me through their process, and catch their error. Then I would give them hints from their prior learning to help them get started” (Howard Stenzel Interview, August 13, 2014). Bill Hollister responded by saying, “I would try to figure out what they were thinking, watch them work and let them explain it. Then I would reteach, using different strategies.” (Bill Hollister Interview, August 15, 2014).

In Table 2 below, we can see the combinations of levels of definitions and examples for each teacher.

**Table 2**

*Teachers' Definitions and Examples by Theme Level*

<table>
<thead>
<tr>
<th>Name</th>
<th>Definition of Formative Assessment</th>
<th>Example of What Teachers Do Next</th>
</tr>
</thead>
<tbody>
<tr>
<td>Julie Jones</td>
<td>1- does student get it?</td>
<td>1- reteach</td>
</tr>
<tr>
<td>Natalie Night</td>
<td>2 -student plus reteach</td>
<td>1- reteach</td>
</tr>
<tr>
<td>McKenzie Orr</td>
<td>1- does student get it?</td>
<td>1- reteach</td>
</tr>
<tr>
<td>Trish Tisby</td>
<td>1- does student get it?</td>
<td>2 - teach differently</td>
</tr>
<tr>
<td>Pat Paulson</td>
<td>2 -student plus reteach</td>
<td>1- reteach</td>
</tr>
<tr>
<td>Howard Stenzel</td>
<td>1- does student get it?</td>
<td>3 - understand student thinking and give hints of prior learning</td>
</tr>
<tr>
<td>Chris Carter</td>
<td>1- does student get it?</td>
<td>1- reteach</td>
</tr>
<tr>
<td>Bill Hollister</td>
<td>1- does student get it?</td>
<td>3 - have student work the problem, to figure out their thinking, reteach using different strategies</td>
</tr>
<tr>
<td>Gwen Stacy</td>
<td>1- does student get it?</td>
<td>1- reteach</td>
</tr>
<tr>
<td>Julian Keller</td>
<td>3 - student plus modify instruction</td>
<td>2 - teach differently</td>
</tr>
<tr>
<td>Cecilia Reyes</td>
<td>1- does student get it?</td>
<td>1- reteach</td>
</tr>
<tr>
<td>Micah Norris</td>
<td>1- does student get it?</td>
<td>1- reteach</td>
</tr>
<tr>
<td>Marie Montee</td>
<td>1- does student get it?</td>
<td>1- reteach</td>
</tr>
</tbody>
</table>
Types of Formative Assessments Teachers Reported Using

The teachers were asked to list the types of formative assessments that they use in class. The group of thirteen teachers gave 81 individual responses to this question. Some of the teachers gave answers that were “type repeats”, meaning that the assessments, while different in name, were of the same type. An example of this would be if a teacher listed thumbs up/thumbs down and asking students to nod yes or no as two assessments, I counted them as type repeats, because while they are different assessments, they both yield a binary response from students. Having students write and having them journal would also be type repeats. The types are based on the responses they elicit from students.

When the assessment type repeats were accounted for, the 81 responses dropped to 59 responses. Those 59 responses were then grouped into the 13 different types shown in Table 3.

Table 3

<table>
<thead>
<tr>
<th>Types of Formative Assessments</th>
<th>Number of Responses</th>
<th>Category of Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Partner/Group Share</td>
<td>10</td>
<td>2 - Procedural</td>
</tr>
<tr>
<td>2 Thumbs up/Thumbs down</td>
<td>10</td>
<td>1 - Scaled</td>
</tr>
<tr>
<td>3 Discussion/Questioning/Conversation</td>
<td>7</td>
<td>2 - Procedural</td>
</tr>
<tr>
<td>4 Observation While Working on Task</td>
<td>7</td>
<td>2 - Procedural</td>
</tr>
<tr>
<td>5 White Boards</td>
<td>7</td>
<td>2 - Procedural</td>
</tr>
<tr>
<td>6 Exit Tickets</td>
<td>4</td>
<td>2 - Procedural</td>
</tr>
<tr>
<td>7 4 Corners</td>
<td>3</td>
<td>1 - Scaled</td>
</tr>
<tr>
<td>8 iPad</td>
<td>3</td>
<td>2 - Procedural</td>
</tr>
<tr>
<td>9 Quiz</td>
<td>3</td>
<td>2 - Procedural</td>
</tr>
<tr>
<td>10 Journal/Writing</td>
<td>2</td>
<td>2 - Procedural</td>
</tr>
<tr>
<td>11 Fist to 5</td>
<td>1</td>
<td>1 - Scaled</td>
</tr>
<tr>
<td>12 Red/Yellow/Green cards</td>
<td>1</td>
<td>1 - Scaled</td>
</tr>
<tr>
<td>13 Can they explain their thinking</td>
<td>1</td>
<td>3 - Conceptual</td>
</tr>
</tbody>
</table>
The 13 types of formative assessments can be grouped into three different categories: scaled (level 1), procedural (level 2), and conceptual (level 3). The three categories are based on the type of response that the assessment elicits from a student. The scaled assessments give the teacher a scaled response from the student and the scales ranged from a thumbs up/thumbs down scale of two to a scale of six using fist to five. With each of the scaled assessments the teacher asks a question of the students and the students self-assess and respond based on the scale being used. As an example, the teacher might ask their students, “tell me how comfortable you are with long division using fist to five, where your fist means that you don’t get it at all and your five fingers out tells me that you are a long division ninja.”

The second category of formative assessment is a procedural assessment. These assessments usually give the teacher information about a student’s ability to do a particular procedure. Seven of the teachers listed white boards as a form of formative assessments. These teachers have notebook sized dry erase boards for each student in their classes, and they will ask students to work a particular problem and then hold their board up for the teacher to see. Three teachers listed iPads as formative assessments and they use the iPads in much the same way the other teachers use the dry erase boards.

The third category of formative assessment is a conceptual assessment. This category of assessment solicits a student’s conceptual understanding of the math being learned. Trish Tisby, who responded, “asking students to explain the why behind procedures”, (Trish Tisby Interview, August 13, 2014) was the only teacher to list formative assessment that explicitly addressed students’ conceptual understanding. It is possible that the assessments listed in the procedural assessment category could be used
to solicit students’ conceptual understandings, however, because these assessments do not necessarily do so, I did not move them to the conceptual category.

When the thirteen teachers listed the types of formative assessments they use, ten of the teachers listed assessments that fell in the first two categories, scaled and procedural assessments. There were no teachers who reported that they only used scaled assessments. Two of the teachers listed formative assessments that only fell into the procedural assessment category. There was one teacher who reported using formative assessments that were both procedural and conceptual. The types of assessments reported by teachers and categories of those assessments are listed below in Table 4.

Teachers were also asked if they imbed formative assessments into their lessons when they plan their lessons. Eight of the teachers indicated that they do not imbed the formative assessments in their lesson plans. Three of the teachers gave responses that indicated it happens sometimes. In response to the question Trish Tisby said, “not as much as I should.” (Trish Tisby Interview, August 13, 2014). Pat Paulson responded by saying, “yes, not all of the time but some.” (Pat Paulson Interview, August 14, 2014) Gwen Stacy responded by saying, “I see the value of it, adding more structure, but I only put it in my lesson plans sometimes."

Two of the teachers, Julie Jones and Julian Keller, indicated that they “usually” put formative assessments into their lesson plans. Teachers’ responses to whether they imbed formative assessments in their lesson plans are also reflected in Table 4.
Table 4

*Teachers' Reported Types of Formative Assessments and Whether Assessments are Imbedded in Lesson Plans*

<table>
<thead>
<tr>
<th>Name</th>
<th>Reported Types of Formative Assessment</th>
<th>Categories of Assessments</th>
<th>Imbed Assessments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Julie Jones</td>
<td>thumbs, 1 to 5, looking at work, discussion questions, partner share, white board</td>
<td>1 - Scaled, 2 - Procedural</td>
<td>Usually</td>
</tr>
<tr>
<td>Natalie Night</td>
<td>thumbs, white boards, table talk, asking question, standup sit down, corners</td>
<td>1 - Scaled, 2 - Procedural</td>
<td>No</td>
</tr>
<tr>
<td>McKenzie Orr</td>
<td>listen, ask questions, turn and talk, monitoring, pull groups, task and listen to them, pair share</td>
<td>2 - Procedural</td>
<td>No</td>
</tr>
<tr>
<td>Trish Tisby</td>
<td>pop quiz, word problem Monday, conversation, can they explain it, teach neighbor</td>
<td>2 - Procedural, 3 - Conceptual</td>
<td>Sometimes</td>
</tr>
<tr>
<td>Pat Paulson</td>
<td>white boards, conversations, in group, independent practice, iPad, thumbs</td>
<td>1 - Scaled, 2 - Procedural</td>
<td>Sometimes</td>
</tr>
<tr>
<td>Howard Stenzel</td>
<td>proximity, open ended questions, thumbs, red/yellow/green cards,</td>
<td>2 - Procedural</td>
<td>No</td>
</tr>
<tr>
<td>Chris Carter</td>
<td>walk around, exit tickets, I-we-you check, thumbs</td>
<td>1 - Scaled, 2 - Procedural</td>
<td>No</td>
</tr>
<tr>
<td>Bill Hollister</td>
<td>white boards, thumbs, partner share, pass the plate, journal, exit tickets, I we you</td>
<td>1 - Scaled, 2 - Procedural</td>
<td>No</td>
</tr>
<tr>
<td>Gwen Stacy</td>
<td>thumbs, 4 corners, think pair share, iPads, shoulder partner, journals, clipboard cruising, white board, exit tickets</td>
<td>1 - Scaled, 2 - Procedural</td>
<td>Sometimes</td>
</tr>
<tr>
<td>Julian Keller</td>
<td>thumbs, white boards, turn and talk, iPad survey, fist to 5</td>
<td>1 - Scaled, 2 - Procedural</td>
<td>Usually</td>
</tr>
<tr>
<td>Cecilia Reyes</td>
<td>short quiz, desk work, thumbs, 4 corners, listening to discussion, writing</td>
<td>1 - Scaled, 2 - Procedural</td>
<td>No</td>
</tr>
<tr>
<td>Micah Norris</td>
<td>thumbs, board competition, calendar time, white boards, observations, shoulder partners, number heads together - listen, short quiz, get or not, think pair share</td>
<td>1 - Scaled, 2 - Procedural</td>
<td>No</td>
</tr>
<tr>
<td>Marie Montee</td>
<td>thumbs, marker boards, top of desk, nod, sticky notes, exit tickets, 4 corners</td>
<td>1 - Scaled, 2 - Procedural</td>
<td>No</td>
</tr>
</tbody>
</table>

**Teachers’ Self Rating of Their Use of Formative Assessment**

Toward the end of each interview, after the discussion of formative assessments, teachers were asked to rank themselves from 1 to 10 (1 being low) on their own use of
formative assessments in their math classrooms. Six of the teachers rated themselves as a 7. Five teachers rated themselves as a 6 while Pat Paulson chose a rating of 9 and Chris Carter chose a rating of 3.

Since none of the teachers gave themselves a 10 on their use of formative assessment, I asked them what they could be doing that would move them towards a rating of 10. All of the teachers responded in the same general way by saying they would be more structured and conscientious about planning for formative assessments and would use them more often. Pat Paulson, who gave a self-rating of 9, responded by saying, “I’m not sure what I would do differently. I feel like you can always improve. Maybe I would add more types.” (Pat Paulson Interview, August 14, 2014). Chris Carter, whose self-rating was 3, said, “I would use them more because I don’t always use them.” (Chris Carter Interview, August 18, 2014). All of the teachers viewed the path to improvement the same way, no matter what rating they gave themselves.

Teachers were also asked what barriers, if any, made improving their use of formative assessments more difficult. The most common answer was time, with eight teachers identifying it as the main barrier to improved use of formative assessments. Chris Carter explained, “When you’re following the pacing guide you have to cover a certain amount of material in the school year. Adding the formative assessments takes time, and throws you off pace.” (Chris Carter Interview, August 18, 2014). Marie Montee and Bill Hollister both identified peer pressure as a barrier. Marie Montee explained, “when I ask students thumbs-up or thumbs-down, some of them will look around before they put their thumb up or down, to see what everyone else is doing. If a
Kid looks around and then gives me a thumbs up, I can’t trust their answer.” (Marie Montee Interview, August 22, 2014).

Two of the teachers, Pat Paulson and Gwen Stacy both indicated that the absence of a resource teacher was a barrier to using formative assessments. Pat Paulson said, “I had a resource teacher last year and it was easy to pull out a kid and go over a problem one on one. It’s harder to do now without the second teacher.” (Pat Paulson Interview, August 14, 2014). Julie Jones and Natalie Night both indicated that they didn’t see any barriers to improving their use.

Teachers’ self-ratings, steps to improvement and perceived barriers to their improvement are listed below in Table 5.

### Table 5
**Teachers' Self Ratings, Steps to Improve, and Perceived Barriers**

<table>
<thead>
<tr>
<th>Name</th>
<th>Self Rating</th>
<th>Steps to Improve</th>
<th>Barriers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Julie Jones</td>
<td>6</td>
<td>used them more often, I don't use them in every lesson</td>
<td>none</td>
</tr>
<tr>
<td>Natalie Night</td>
<td>7</td>
<td>really planning and imbedding them into the lesson</td>
<td>none</td>
</tr>
<tr>
<td>McKenzie Orr</td>
<td>7</td>
<td>want to check more often</td>
<td>time</td>
</tr>
<tr>
<td>Trish Tisby</td>
<td>6</td>
<td>more conscientious use and implementation of them</td>
<td>time</td>
</tr>
<tr>
<td>Pat Paulson</td>
<td>9</td>
<td>not sure, feel like can always improve, add more types</td>
<td>absence of resource teacher</td>
</tr>
<tr>
<td>Howard Stenzel</td>
<td>7</td>
<td>would plan them better and use more often</td>
<td>time</td>
</tr>
<tr>
<td>Chris Carter</td>
<td>3</td>
<td>use more often, don't always do them,</td>
<td>time</td>
</tr>
<tr>
<td>Bill Hollister</td>
<td>6</td>
<td>need to add variety, use more often,</td>
<td>honesty of kids, peer pressure,</td>
</tr>
<tr>
<td>Gwen Stacy</td>
<td>6</td>
<td>consciously thinking about opportunities for FA in the planning, not the natural</td>
<td>absence of resource teacher,</td>
</tr>
<tr>
<td>Julian Keller</td>
<td>7</td>
<td>remembering to write down and get back to student, more structure</td>
<td>time</td>
</tr>
<tr>
<td>Cecilia Reyes</td>
<td>7</td>
<td>better planning to assess more</td>
<td>time, paperwork</td>
</tr>
<tr>
<td>Micah Norris</td>
<td>6</td>
<td>would be assessing more often</td>
<td>time</td>
</tr>
<tr>
<td>Marie Montee</td>
<td>7</td>
<td>want to do more often</td>
<td>peer pressure</td>
</tr>
</tbody>
</table>
Chapter Summary

All of the teachers expressed an understanding of what formative assessment was, and also expressed a belief that there was value to using formative assessments in class. All of the teachers were able to give a definition of formative assessment, although there was some variation in the definitions. The definitions ranged from simply seeing if a student got it or not, to trying to understanding student thinking.

When asked to list specific types of formative assessments, the answers ranged from a dichotomous thumbs-up/thumbs down to asking a student to explain their thinking. Most of the identified types of formative assessments were procedural in nature.

When rating themselves on their effective use of formative assessments, teachers averaged a 6.45 on a 10-point scale (1 being low), with only one teacher self-rating below the median at a 3. When asked what they could do to improve, most agreed that improvement would come through being more intentional about the use of formative assessments through planning and using formative assessments more often. The main barrier to improvement that teachers perceived was time in the classroom.

The teacher interview data creates a picture of how teachers perceive their use of formative assessments in their math classrooms. In the next chapter data collected from teachers actual use of formative assessments in their classrooms will be examined, and ultimately their perceptions will be compared to their actual use.
CHAPTER 4: TEACHERS’ ACTUAL IMPLEMENTATION OF FORMATIVE ASSESSMENT

Observing teachers’ actual implementation of formative assessments gives us a better insight into how teachers understand and use formative assessments. While observing teachers, I was looking for how the teachers’ implemented the formative assessments. This included the frequency of their use during a class period, the types of formative assessments that the teachers used, and the data that was actually gathered from the use of the formative assessments. I also sought to document how the teacher used the data from the formative assessments.

Frequency of Use of Formative Assessments

The current literature on formative assessment does not specifically state how often formative assessments should be used in the classroom, or even how to measure frequency. The National Council of Teachers of Mathematics state in their position statement on the role of formative assessments in math education that it “strongly endorses the integration of formative assessment strategies into daily instruction.” (NCTM, 2013) The National Council of Supervisors of Mathematics position on formative assessment states, “formative assessment needs to be intentionally and systematically integrated into classroom instruction at every grade level.” (NCSM, 2014).

Because of the variety of formative assessment options that teachers have, determining frequency is not as straightforward as it may initially seem. A thumbs up/thumbs down type assessment may only take 5 seconds. If the teacher gives students a real world, multi-step question to work on, and walks around looking at their work and asking about their thinking, it may take 20 minutes. Because of this, I decided that
instead of trying to count individual assessments, it makes more sense to quantify the
amount of time a teacher spends in an assessment mode. I am defining assessment mode
as any time when the teacher is specifically looking for student feedback on a question or
task that is related to the lesson objective.

During the classroom observations of teachers I tracked the start and stop times
when teachers were in assessment mode during the 60 minute lessons. I averaged the
times of the two observations for each teacher and rounded the average to the nearest
minute. There were two teachers, Trish Tisby and Bill Hollister, who each had a full
class period when the students didn’t do math. Because of this, for these two teachers I
only used the assessment time for the class where math was being taught. Table 6 below
lists the average time in assessment mode for each teacher.

**Table 6**  
*Teachers' Average Time In Assessment Mode Per Lesson*

<table>
<thead>
<tr>
<th>Name</th>
<th>Time in Minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Julie Jones</td>
<td>9</td>
</tr>
<tr>
<td>Natalie Night</td>
<td>23</td>
</tr>
<tr>
<td>McKenzie Orr</td>
<td>18</td>
</tr>
<tr>
<td>Trish Tisby</td>
<td>30*</td>
</tr>
<tr>
<td>Pat Paulson</td>
<td>37</td>
</tr>
<tr>
<td>Howard Stenzel</td>
<td>15</td>
</tr>
<tr>
<td>Chris Carter</td>
<td>23</td>
</tr>
<tr>
<td>Bill Hollister</td>
<td>25*</td>
</tr>
<tr>
<td>Gwen Stacy</td>
<td>8</td>
</tr>
<tr>
<td>Julian Keller</td>
<td>35</td>
</tr>
<tr>
<td>Cecilia Reyes</td>
<td>10</td>
</tr>
<tr>
<td>Micah Norris</td>
<td>15</td>
</tr>
<tr>
<td>Marie Montee</td>
<td>9</td>
</tr>
</tbody>
</table>

* Denotes time from one class, because the full period of the second class was taken up by non-math tasks.*
Types of Formative Assessments Actually Used

All of the teachers used formative assessments in each of their math classes with the exception of two class periods. Trish Tisby and Bill Hollister each had one class where the lesson for that day was used to set up the math for the next day’s lesson. In Trish Tisby’s class, the lesson was to redesign the classroom, picking out new furniture and supplies for the class. The lesson objective was “I can add two digit numbers with regrouping.” The students spent the whole period working in small groups picking supplies and furniture out of catalogues and writing them down with their prices. In the lesson for the following day, the students would be asked to start adding up the cost of their design choices in the context of a budget. Bill Hollister’s class was doing much the same, planning a weeks worth of meals for a family of four based on prices from grocery store fliers. On the first day of the lesson the students were planning meals, and writing the prices down, but did not start their number manipulations yet.

Ten of the teachers used scaled (level one) assessments. The level one assessments that were used included thumbs up/thumbs down and fist to five, which were two of the four scaled assessments listed during the interviews. None of the teachers used the two other scaled assessments (4-corners and red/yellow/green cards) that were listed during the interviews but several teachers used scaled assessments that weren’t listed during the interviews. The new scaled assessments were “clap if you understand” and “stand up if you understand.” Trish Tisby also asked the question, “Does that make sense?” to the whole class, and most of the students responded out loud with “yes”. All of the teachers who used scaled assessments also used procedural (level 2) assessments, so there were no teachers who only used scaled assessments.
All of the teachers used procedural assessments. They used pair-share, gave students a task and walk around observing, white boards, exit tickets, quizzes and discussion/questioning. During my observations I did not see any teachers use journaling or iPads as formative assessments. The whiteboards were used with short math problems that students would work out on the whiteboard. Julie Jones used the whiteboards in a lesson on rounding by calling out a number like 122, and having students round to the nearest 10. Marie Montee used this process to have students convert fractions to decimals, calling out decimals like .52 and .673.

Seven of the teachers used some form of pair-share during their lessons. Pair sharing is distinct from students working together as partners in that, with pair-share, each student comes up with an individual answer and then shares it with someone else. Each time it was used, it was used with a specific and short question. Gwen Stacy asked her class, “why is rounding important?” and then asked students to, “turn to your shoulder partner and talk about why rounding is important.” McKenzie Orr was working on rounding with her class also. She asked, “does everyone know the rule on rounding? Talk to you face partner and tell them the rule about rounding.” When teachers used the pair-share technique, all of the students discussed the questions at one time, and the teachers did not walk around to try and listen to student responses. Five of the seven teachers, after using pair-share would ask a student or two to share with the whole group what they talked about in their pair.

Micah Norris used an exit ticket that had a single division problem (1831 divided by 51) that asked students to use partial quotients to divide. The teacher explained,
“some of the students seemed to have trouble with it during class so I’ll grade the exit
tickets tonight so I know which students to pull out to review tomorrow.”

All thirteen teachers gave students a math task and walked around looking at their
work. The tasks took several forms. Trish Tisby and Natalie Night gave student dice to
roll and use the numbers to connect repeated addition to multiplication. If a student
rolled a 5 and a 6, they had to draw 5 groups of 6 things. The other teachers either used
worksheets, or put a series of problems up on the SMART Board for students to work
through while they walked around and looked at work.

While all of the teachers used procedural assessments, there were two teachers
who only used procedural assessments. McKenzie Orr and Howard Stenzel did not use
any scaled assessments or any conceptual (level 3) assessments. Bill Hollister was the
only teacher who used conceptual assessments, and those assessments came in the form
of follow up questions to a task. Bill Hollister used the follow up questions to understand
student thinking (conceptual questions), and used them with regularity. In one instance
students were creating division equations from a word problem. The task they were
given was the following word problem: “Two girls were selling Girl Scout Cookies, and
together they went to 18 houses. If the two girls went to the same number of houses, how
many houses did each girl go to?” Bill Hollister had the following exchange with a
student:

Teacher: Ryan, what did you get?
Ryan: Each girl went to 9 houses.
Teacher: How did you get that?
Ryan: \(18 - 9 = 9\)
Teacher: Ryan, tell me what you’re thinking.
Ryan: There were 18 houses so if one girl went to nine, then the
other girl went to nine because 9 plus 9 is 18.
Teacher: Now remember, we’re supposed to make a division equation.
Ryan: (pause) never mind
Teacher: No, don’t never mind me, who can help Ryan out… Sam?
Sam: You have two 9s, so 2 times nine is 18, and 18 divided by 2 is 9.

After this exchange the teacher asked, “does anyone else want to share their thinking?”

Another student, Tom, responded, saying:

Tom: I have that they went to 36 houses.
Teacher: How did you get that?
Tom: Right here it says each girl went to 18 houses, so 18 plus 18 is 36.
( at this point a number of students start to try and correct Tom)
Teacher: Everyone hold on… explain what you’re thinking Tom.
Tom: The problem said there were two girls and they went to 18 houses so 18 plus 18 is 36, not just 9.
Teacher: So Tom is saying 18 plus 18 is 36. Talk in your group, and lets see if we know what Tom is doing here.
(students in Tom’s group talk for about 30 seconds)
Teacher: Ok, did we solve the problem?
Student: Tom didn’t understand the question right, and didn’t know that the 18 meant the number of houses all together.
Teacher: Ok, does that make sense now Tom?
Tom: Yes.

Later in the same lesson, the students were still doing division, but now with remainders. Bill Hollister gave student the task, “You have 13 marbles that you are going to divide between 4 friends. How many does each friend get, and how many are left over?” The students were working in small groups. And the teacher asked the red group to report out their answer. Their answer was that each person gets 3 marbles with one left over. All of the other groups agreed, except two students, who disagreed. Bill Hollister asked the two students to explain their thinking.

Teacher: So John and Tim, you need to argue your point, why do you disagree?
John: We made it even, so everyone gets 3 and ¼ marbles.
(some students begin to object –“you can’t split a marble…”)

Teacher: Now hold on… lets listen to what they say… lets see if it works out. You both did the same thing, so explain how to draw the picture.

John: You make 4 circles and you still put 3 marbles in each circle, and then next to the 3 you put a 1 and a slash and a 4.

Teacher: So Tim, you’re good with what John says?

Tim: uhm-hum

Teacher: Let’s look at the math separate from a physical marble… If we have a quarter of marble in this group, and this group and this group and this group, does it add up to 1 marble?

Students: Yes

Teacher: When John was explaining what they did, I started to think about a dollar bill. Because, how many quarters are in a dollar bill?

Students: Four

Teacher: If we were measuring this, I would say they were able to defend their reasoning. Now I’m not saying that this is always true. There will be times when you won’t be able to divide something into fractions, like if they were dividing puppies.

Bill Hollister was the single teacher who used conceptual questioning. The conceptual questions were part of a larger formative assessment process, beginning with the task, and that being the case, the full assessment would be seen as being a conceptual assessment. Bill Hollister asked follow up questions that allowed students to give a variety of thinking responses (as opposed to students all repeating the same process). His continuation of the formative assessments in the form of follow up questions allowed him to elicit data regarding student thinking. Next I examine the data produced from the formative assessments of the other teachers.

**Data From the Formative Assessments Tasks**

During the interviews I asked teachers about the types of formative assessments they used and those assessments were put into three categories. During the observations I tracked the types of formative assessments that each teacher used. When I reviewed the data from how teachers actually used formative assessment, it was clear that the vast
majority (> 95%) of teachers’ time in assessment mode was used by the assessments categorized in Table 3 as procedural, and specifically assessment type 4 – Observation while working on task. For every teacher, this was dominant type of formative assessment.

Because giving a task and observing students as they worked was the dominant mode of formative assessment, it was important to do some analysis on the initial tasks that teachers were using for this assessment. After reviewing all of the teachers’ tasks together, the themes that emerged which separated the tasks were based on the type of data that the task itself elicited from students. The distinction was whether the task asked students for a specific process or whether the task allowed students to approach problems in any way that made sense to them. I labeled this distinction in tasks as closed tasks and open tasks.

Eight of the teachers gave only closed tasks; tasks that asked students to do a specific type of math using a specific process. For example, Natalie Night and Trish Tisby both had students make arrays to show multiplication. Natalie Night used a worksheet and Trish Tisby had students roll dice to come up with the factors for the multiplication. Julie Jones, McKenzie Orr and Gwen Stacy each gave their students a series of numbers to round to the nearest tens or hundreds place. For each of these eight teachers the initial data that these tasks solicited from students only indicated whether or not the students could do a particular math process.

It is important to note that a formative assessment is defined as the tools and processes used to understand student thinking. This means that if the teacher gives a task and watches the students work through the task, that is a formative assessment. If a
teacher gives a task, watches students work through the task, and asks students questions about the task, this is also a formative assessment.

Each of the eight teachers above who used closed tasks also used follow-up questions with students around the given task. All of the follow up questions that these eight teachers used were of a procedural nature, such as, “what is the next step.” So for all of the teachers who used closed tasks, the questioning that followed was procedural. Table 7 below lists whether teachers used open or closed tasks, and whether the teachers’ subsequent questioning was procedural or conceptual.

**Table 7**

<table>
<thead>
<tr>
<th>Name</th>
<th>Tasks</th>
<th>Questioning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Julie Jones</td>
<td>Closed Tasks</td>
<td>Procedural</td>
</tr>
<tr>
<td>Natalie Night</td>
<td>Closed Tasks</td>
<td>Procedural</td>
</tr>
<tr>
<td>McKenzie Orr</td>
<td>Closed Tasks</td>
<td>Procedural</td>
</tr>
<tr>
<td>Trish Tisby</td>
<td>Open Tasks</td>
<td>Procedural</td>
</tr>
<tr>
<td>Pat Paulson</td>
<td>Open Tasks</td>
<td>Procedural</td>
</tr>
<tr>
<td>Howard Stenzel</td>
<td>Closed Tasks</td>
<td>Procedural</td>
</tr>
<tr>
<td>Chris Carter</td>
<td>Open Tasks</td>
<td>Procedural</td>
</tr>
<tr>
<td>Bill Hollister</td>
<td>Open Tasks</td>
<td>Conceptual</td>
</tr>
<tr>
<td>Gwen Stacy</td>
<td>Closed Tasks</td>
<td>Procedural</td>
</tr>
<tr>
<td>Julian Keller</td>
<td>Open Tasks</td>
<td>Procedural</td>
</tr>
<tr>
<td>Cecilia Reyes</td>
<td>Closed Tasks</td>
<td>Procedural</td>
</tr>
<tr>
<td>Micah Norris</td>
<td>Closed Tasks</td>
<td>Procedural</td>
</tr>
<tr>
<td>Marie Montee</td>
<td>Closed Tasks</td>
<td>Procedural</td>
</tr>
</tbody>
</table>

Five of the teachers used open tasks. Pat Paulson asked students, “How many cars would we need to take all of the 3rd graders on a field trip?” as part of a multiplication lesson. Chris Carter asked students to figure out “how many days old are you?” Bill Hollister had students dividing marbles as discussed above, and Julian Keller had student distribute a list of school supplies to each of the students in the 4th grade as part of a lesson that applied division in real world problems. Trish Tisby dedicated a
whole class period to the task of having students redesign the classroom. The students were picking out new furniture and supplies and pricing them.

In each of these instances students were not given a specific process they had to follow to solve the problems. Because these were open tasks, they allowed students to solve using whatever method made sense to them, giving these tasks the potential to give teachers more data about student thinking than a closed task would have. For four of these five teachers, their follow up questions were of a procedural nature. The exception was Bill Hollister, who asked conceptual questions.

**How Teachers Used/Responded to the Data From Formative Assessments**

Each of the ten teachers who used scaled assessments responded to the data they received from the assessment in the same way. Each of the teachers just moved on with the lesson. There was no instance where a teacher asked any follow up questions from a scaled assessment. None of the teachers went back and reviewed any information or pulled any students aside for individual attention following a scaled assessment.

All of the teachers used procedural assessments (specifically, giving a task and watching students work through it), and each of the teachers responded in some way to the data they gathered from students. The eight teachers who used closed tasks in their procedural assessments responded to student work in much the same way. Julie Jones asked students to round a series of numbers to the nearest 10, and when looking at one student’s work, had the following exchange:

Teacher: 122 – look next door, what do you find?  
Student: 120 and 130…?  
Teacher: So I do…? (pausing to invite the student to answer)  
Student: (student stares at the paper silently)  
Teacher: Is it closer to 120 or 130?  
Student: (after a pause ) 120?
Teacher: Yes, so you round down.

McKenzie Orr had a similar exchange in a rounding lesson:

Teacher: What do you think 25 is rounded to?
Student: 50?
Teacher: Think about counting by 10s. What is the 10 before 25?
Student: 20…
Teacher: What is the 10 after 25?
Student: 30…
Teacher: So what do you do?
Student: Round up?
Teacher: Right. It rounds up to 30.

Marie Montee had students converting decimals to fractions, and students were given the decimal number .52 to convert to a fraction. While walking around looking at student work Marie Montee had the following exchange:

Teacher: You forgot the last step. What else do you have to do?
Student: (stares blankly at her work)
Teacher: Simplify. Watch your denominator. Find the number that divides into both of them.

The most common teacher response was to ask the student “what’s the next step?”

While doing long division Cecilia Reyes looked at a student’s work and had the following exchange:

Teacher: We’re going to divide, what’s the next step?
Student: Subtract. (writes something on his paper)
Teacher: What’s the last step?
Student: Bring down. (writes some more)
Teacher: (pointing at part of the problem) How many times does this go into this?
Student: (student writes some more)
Teacher: What’s the next step? Subtract. Go ahead and subtract.

In some cases the teacher indicated to the student that their answer was wrong and then told the student how to do the problem. Micah Norris was teaching division using
partial quotients, and the students had the problem $1831 \div 51$. The following exchange occurred with one student:

Teacher: No, that’s not how you do it. What’s 10 times 51?
Student: 510…
Teacher: Subtract 510 from 1831, and put the 10 over here. Can you subtract another 510 from the remainder?
Student: Yes…
Teacher: Go ahead and subtract it, and put another 10 over here. Can you do it again?
Student: Uhm…. Yes…
Teacher: Ok, do it again and put another 10 over here. Now how many times will 51 go into that?
Student: (working on paper a bit) 5…
Teacher: Ok, put your 5 over here and subtract these. So how many times did 51 go into 1831?
Student: (adding the 10s and the 5) uhm… 35

The eight teachers who used closed tasks responded to students who were stuck or had incorrect answers by walking the students through the specific procedures. The teachers did not ask any questions to have students explain how they were thinking about the problem, or what they did or did not understand about the process they were learning. If a student did the problem correctly, the teachers would give them some affirmation, such as, “that’s right”, or “good job,” but did not ask those students any questions.

There were five teachers who used open tasks, which could allow for a variety of processes from students. When each of these teachers gave the task to the students they did not indicate a specific process in the instructions. Pat Paulson asked how many cars would be needed to take all the 3rd graders on a field trip. The students worked in groups, and each group had to decide what type of vehicle they were going to use (sedan, SUV, truck, etc.). All the groups realized they needed to know how many 3rd graders there were, and they were told 157. At this point each group of students had to figure out how many vehicles they would need for 157 students.
Pat Paulson moved around the room looking at student work, and began talking to students saying things like: “3 x 4 means 3 groups of 4, always, always. So if the SUV’s hold 6 students then 6 is the size of your group and you have to find how many groups of 6 there will be,” and “the number of students in your car is the size of your group. When we set up multiplication which number is the size of the group?” Even though Pat Paulson did not indicate a specific process to be used in the initial instructions for the task, it was clear from the responses to student work that the teacher had a specific method in mind and was steering students to use that method.

Pat Paulson did not ask students to explain their processes or thinking. Once it was clear that a student was not using the preferred method, the student was guided in the direction of the preferred method. While this was an open task, because Pat Paulson didn’t ask any further questions that addressed students’ conceptual understanding of multiplication and division, the assessment remained a procedural assessment.

Chris Carter and Julian Keller also had open tasks that did not specify a particular process, but once the students started working, both teachers moved students to a single process, without asking any further questions to solicit students’ conceptual understandings of the math in the tasks. Chris Carter asked students “how many days old are you?” and walk around as students worked, repeatedly telling students “you have to start with the number of years you are and multiply that by the number of days in a year.” Several students had used other methods to answer the question, but were made to go back and use this process.

Bill Hollister was the only teacher who used conceptual questions as follow-up to the open task. He started with an open task, (dividing marbles between friends) and
allowed students to work through the task in the way that made the most sense to them. When students had different answers, they had to explain and justify their reasoning, and Bill Hollister allowed the students to clear up misconceptions with each other.

Bill Hollister’s was also the only observed lesson where the teacher modified instruction during class based on the students. When the student brought up dividing the last marble into quarter pieces, Bill Hollister allowed the discussion of fractions to happen, even though it was not the math being reviewed that day. Bill Hollister allowed the discussion to occur and supplemented the thinking of the student by adding a short discussion about money (quarters) to further clarify the use of fractions even though it was not part of his original lesson.

Part of my definition of formative assessment is, “engaging students to be active in their own academic success.” The traditional model of education has been that teachers impart knowledge to students who are the passive recipients of that knowledge. Ideally, part of using formative is helping students become active participants in their own learning. By allowing students to figure things out, to construct, test, explain and justify knowledge, students become active in their own learning.

Bill Hollister encouraged student-to-student interaction. The students in his class were encouraged to present their own thinking, not just a repetition of a process that was given to them by the teacher. Students were asked if they disagreed with each other, and encouraged to justify their reasoning when they disagreed. Students were encouraged to help clear up misconceptions for each other. It is this type of interaction that allows students to become active in their own learning. Bill Hollister’s class was the only class where I observed this type of interaction between students.
Chapter Summary

All of the teachers in the study used formative assessments during their lessons. The teachers were generally accurate when comparing the types of assessments that teachers reported using in their interviews to the types of assessments that they actually used in class. Even though many assessments were listed and used, the overwhelming majority of teachers’ time in assessment mode was used on assessment type 4; giving a task and observing students’ work.

The teachers’ average time in assessment mode during class ranged from a low of 8 minutes to a high of 37 minutes during a 60-minute class period.

The tasks that teachers used were usually closed tasks, with eight of the teachers using these types of task to gauge students’ ability to use a procedure. Five of the teachers used open tasks, with four of these teachers using the open tasks followed by procedural questioning and only one using the open tasks with conceptual questioning.

Looking at teachers’ reported understanding and use of formative assessment along with the data from their actual use we can create a Depth of Implementation Framework to assist in the understanding of how teachers are using formative assessments in the classes.
CHAPTER 5: DEPTH OF IMPLEMENTATION FRAMEWORK

When new educational methodologies are developed, the developers often create a framework and process for evaluating the implementation of that methodology. Doing this helps everyone who is trying to implement and support the new methodology. For methodologies like formative assessment, that developed organically in classrooms, and were studied after the fact, there is no universal handbook that tells an educator how to implement, support and assess the use of formative assessment. In this chapter, I examine teachers’ understanding and use of formative assessment to create a framework that characterizes how deeply a teacher is implementing formative assessment.

The starting point of the framework is a definition of formative assessment. Because there is no consensus in the literature on the definition formative assessment or the necessary elements of formative assessment implementation, I developed the definition of formative assessment below based on the practices that the research supported as being effective in the classroom.

Formative assessment is made up of the tools and processes used during teaching and learning to raise student achievement by gathering information to understand student thinking, providing helpful student feedback, modifying our instruction based on that information, and engaging students to be active in their own academic success. Information may be gathered through a formal written instrument or informally through planned activities and discussion or through unplanned opportunities such as discussion between and amongst students.

This definition is helpful as a starting point for looking at teachers’ use and understanding of formative assessment, however, it does not tell us how to gauge the use of formative assessment. Within each element of the definition there are a variety of ways that teachers could implement. There are also a variety of combinations of parts of
the definition that teachers could implement. Instead of guessing at what the variety of implementation patterns should look like, I use the data to create a theoretical framework. The themes that emerge from the teacher data give a better understanding of teacher’s actual use of formative assessments, and arranging those themes in a Depth of Implementation Framework allows for a better understanding of the depth to which teachers implement formative assessment in their classrooms.

**Depth Of Implementation Framework**

The Depth of Implementation Framework looks at the data in three domains: Intentionality of Integration, Inquiry and Interaction, and Modification of Instruction. Within each domain are the elements derived from the data and discussed in Chapters 3 and 4, which will determine at what level the teacher is implementing formative assessment. The elements within each domain are:

**Domain 1: Intentionality of Integration**
- teacher definition
- teacher examples
- planning
- barriers

**Domain 2: Inquiry and Interaction**
- tasks
- questioning
- time in assessment mode

**Domain 3: Modification of Instruction**
- actual use
- promotion of active learning

The three domains are a progression. Domain 1, Intentionality of Integration, is focused on what is happening before the lesson – how a teacher thinks of and plans for formative assessment. Domain 2, Inquiry and Interaction, is what is happening during
the lesson. This second domain is based on the fourth type of formative assessment, observation of students while they work on a task. Because more than 95% of teachers’ time in assessment mode was used on the 4th type of assessment, it shaped the second domain. Domain 3, Modification of Instruction, looks at how, if at all, teachers change their instruction based on data from students. It also looks at what, if anything, teachers do to engage students to be active in their own learning.

In Chapters 3 and 4 I reviewed the data for each of the elements within the domains, and used the emerging themes to stratify the data. Depth of implementation is determined by looking at all of the stratified data together, and determining in which of three levels each teacher’s implementation of formative assessment belongs: superficial implementation, developing implementation or deep implementation.

**Domain 1 - Intentionality of Integration**

This domain contains four self-reported elements; teachers’ definitions of formative assessment, teacher examples of their actions in response to student data, whether teachers plan their assessments, and teachers’ perceived barriers to implementation. When viewed together, these four elements paint a picture of how intentional teachers are being with their implementation of formative assessment.

The levels of stratification for each of the four elements is as follows:

**Teacher Definitions:**
- level 1 - check to see if a student gets it
- level 2 - check to see if a student gets it plus teacher reteaches
- level 3 - check to see if a student gets it plus a modification of instruction

**Teacher Examples:**
- level 1 - reteach the material
- level 2 - reteach the material differently (modify instruction)
- level 3 - understand student thinking then modify instruction
Teacher Plans Formative Assessment:
   level 1 - no
   level 2 - sometimes
   level 3 - usually

Teachers’ Perceived Barriers:
   - none
   - time
   - peer pressure
   - no resource teacher

The data under Teachers’ Perceived Barriers is not stratified, however it does give us information about how the teachers view formative assessment.

**Domain 2 - Inquiry and Interaction**

This domain looks at the beginning of the formative assessment process, where the teacher gives students a task or question to elicit student data, and the questions that teachers used in response to the data they received from students. This domain also captures the amount of time each teacher spent in assessment mode. The elements in this domain are:

- **Tasks Used to Solicit Student Data**
  - open tasks
  - closed tasks

- **Questioning Used to Follow-Up Tasks**
  - procedural
  - conceptual

- **Time Spent in Assessment Mode**
  - above the average (19.8)
  - below the average

For the element of time, teachers were actually timed during their classroom observations, and the actual time in assessment mode is used. The average time across all observed teachers was 19.8 minutes per class period.
Domain 3 - Modification of Instruction

In this domain, I looked at teachers’ actions to see if they changed in some significant way what or how they were instructing based upon student data. Here, the data indicates either yes there was a modification or no there was not a modification.

I also looked for teacher actions that allowed or encouraged students to be active in their own academic success. The data here also indicates either a yes or a no.

Data from all three domains are used to determine teachers’ levels of implementation. The second and third domains, however, are given more weight because the data from these domains capture what actually happen versus what a teacher was thinking. I also did not use a strict numeral weighting to determine the depth of implementation, but rather, looked for the divisions in the data that naturally emerged. The teachers did, in fact, stratify into three distinct groups that were easily identified.

Now that the data has been presented and organized it is apparent that there are a lot of moving parts. It is easier to understand the data when it is visually presented. All of the data being used for Depth of Implementation analysis is presented in Table 8: Depth of Implementation Data on the following page.
<table>
<thead>
<tr>
<th>Name</th>
<th>Definition</th>
<th>Examples</th>
<th>Planning</th>
<th>Barriers</th>
<th>Tasks</th>
<th>Questioning</th>
<th>Time</th>
<th>Modified</th>
<th>Active</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bill Hollister</td>
<td>1- does student get it?</td>
<td></td>
<td>1- no</td>
<td>peer pressure</td>
<td>Open</td>
<td>Conceptual</td>
<td>25</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Julian Keller</td>
<td>3. student plus modify instruction</td>
<td>2 - teach differently</td>
<td>3 - usually</td>
<td>time</td>
<td>Open</td>
<td>Procedural</td>
<td>35</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Pat Paulson</td>
<td>2 - student plus reteach</td>
<td>1 - reteach</td>
<td>2-sometimes</td>
<td>no resource teacher</td>
<td>Open</td>
<td>Procedural</td>
<td>37</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Trish Tilby</td>
<td>1- does student get it?</td>
<td></td>
<td>2 - teach differently</td>
<td>2-sometimes</td>
<td>time</td>
<td>Open</td>
<td>Procedural</td>
<td>30</td>
<td>No</td>
</tr>
<tr>
<td>Chris Carter</td>
<td>1- does student get it?</td>
<td>1 - reteach</td>
<td>1- no</td>
<td>time</td>
<td>Open</td>
<td>Procedural</td>
<td>23</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Natalie Night</td>
<td>2 - student plus reteach</td>
<td>1 - reteach</td>
<td>1- no</td>
<td>none</td>
<td>Closed</td>
<td>Procedural</td>
<td>23</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Howard Stenzel</td>
<td>1- does student get it?</td>
<td></td>
<td>3. understand student thinking and give hints of prior learning</td>
<td>1- no</td>
<td>time</td>
<td>Closed</td>
<td>Procedural</td>
<td>15</td>
<td>No</td>
</tr>
<tr>
<td>McKenzie Orr</td>
<td>1- does student get it?</td>
<td>1 - reteach</td>
<td>1- no</td>
<td>time</td>
<td>Closed</td>
<td>Procedural</td>
<td>18</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Micah Norris</td>
<td>1- does student get it?</td>
<td>1 - reteach</td>
<td>1- no</td>
<td>time</td>
<td>Closed</td>
<td>Procedural</td>
<td>15</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Cecilia Reyes</td>
<td>1- does student get it?</td>
<td>1 - reteach</td>
<td>1- no</td>
<td>time</td>
<td>Closed</td>
<td>Procedural</td>
<td>10</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Julie Jones</td>
<td>1- does student get it?</td>
<td>1- reteach</td>
<td>3 - usually</td>
<td>none</td>
<td>Closed</td>
<td>Procedural</td>
<td>9</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Marie Montee</td>
<td>1- does student get it?</td>
<td>1 - reteach</td>
<td>1- no</td>
<td>peer pressure</td>
<td>Closed</td>
<td>Procedural</td>
<td>9</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Gwen Stacy</td>
<td>1- does student get it?</td>
<td>1- reteach</td>
<td>2-sometimes</td>
<td>no resource teacher</td>
<td>Closed</td>
<td>Procedural</td>
<td>8</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
Deep Implementation
For a teacher to be classified as having deep implementation of formative assessment, they would have to consistently exhibit high or positive indicators across the second and third domains (Inquiry and Interaction, and Modification of Instruction) because these are the domains that reflect what actually happened during instruction. Only one teacher, Bill Hollister, registered positive indicators in the third domain, Modification of Instruction. Bill Hollister also had positive indicators across all three elements of the second domain, Inquiry and Interaction. Bill Hollister’s data clearly classified him as having deep implementation.

Bill Hollister’s indicators were positive across all of the elements of the second and third domains, so it was easy to classify his implementation of formative assessment as deep. It is conceivable that a teacher could be classified as having deep implementation and not have every single indicator in the second and third domains be positive or high. In this study there were no teachers whose data indicated they were close to the line between deep implementation and developing implementation. In the absence of this data, I chose not to arbitrarily decide which combinations of high and low indicators would constitute the line between deep implementation and developing implementation.

Intentionality of Integration
While this first domain, is not determinative of a teacher’s implementation of formative assessment being classified as deep, it does help give a sense of a whether there is continuity between the teacher’s understanding and the teacher’s use of formative assessment. This domain can also be used to determine where within the classification of deep a teacher’s use of formative assessment falls. Not all deep implementation is the
same. There may be some teachers on the low end of deep and some teachers on the high end of deep.

Bill Hollister’s indicators in this domain were not all high and positive. All of these indicators come from self-reported data, and Bill Hollister did not consistently articulate an understanding of formative assessment that indicated an understanding of deep implementation.

**Definition and Example:**

Bill Hollister gave a level 1 definition of formative assessment that just sought to determine whether or not a student “got it.” When asked to give an example of what happens when a student doesn’t get it, Bill Hollister said, “I usually pull the student aside at some point and reteach the material using different strategies. I have them work the problem, to figure out what they are thinking. I watch them work the problem, let them explain it.” (Bill Hollister Interview, August 15, 2014.)

While Bill Hollister’s definition was level 1, the example he gave of what happens next demonstrated the need to understand student thinking, and recognized that the teacher needed to modify instruction, which put Bill Hollister’s example response at level 3.

**Planning:**

Bill Hollister indicated in his interview that he did not normally plan formative assessments as part of the lesson planning process. This is a low level response and would normally indicate that a teacher is not as intentional in using formative assessments as he could be.
Barriers:

Bill Hollister indicated that peer pressure was a barrier that had a negative effect on formative assessments. Specifically, in certain instances some students would not admit that they did not understand something because their peers seemed to be getting it. By describing peer pressure as a barrier to the effective use of formative assessment, Bill Hollister seems to indicate that he thinks of formative assessments as the low level dichotomous check for understanding. An example from one of the teachers was a thumbs up or down assessment, where a student can look around at his peers and decide not to put his thumb down if everyone else had their thumbs up.

The elements in Bill Hollister’s Intentionality domain are a bit of a mixed bag, ranging from low to high. Looking at Bill Hollister’s other two domains, however, made it clear that what is actually happening in the classroom is the consistent use of formative assessment with deep implementation.

Inquiry and Interaction

In the second domain, Inquiry and Interaction, Bill Hollister used open tasks as part of his formative assessments, such as the Girl Scouts going to houses questions and the dividing marbles question, which allowed students to approach the math in a variety of ways. Bill Hollister also used conceptual assessment questions as follow-up to the tasks, such as, “tell me what you’re thinking”, which allowed students to explain their thinking, instead of forcing students to focusing on a specific process. For these two elements of this domain Bill Hollister exhibited positive indicators.
Bill Hollister also spent 25 minutes of a 60-minute lesson in assessment mode. This was above the median of 19.8 minutes, and tends to exhibit the teacher’s commitment in using formative assessment.

*Modification of Instruction*

In the third domain, Modification of Instruction, Bill Hollister modified instruction in two significant ways. First, he allowed a discussion to go in a completely different direction than planned. When discussing the division of marbles question, a student introduced fractions, which was not a part of the lesson, and was a completely new math topic. Instead of shutting it down, Bill Hollister allowed the discussion to go on, and then augmented the concept by introducing money and dividing dollars by quarters. Bill Hollister allowed a student’s thinking to change the course of instruction, and used it to enhance student thinking about the topic at hand.

Bill Hollister modified instruction in a second important way by encouraging students to be active learners. He prompted the students to not only explain what they were thinking, but to disagree with each other, and then debate the question. She allowed students to interact with each other in ways that lead students to teach each other, instead of knowledge always coming from the teacher. By encouraging students to do this, Bill Hollister engaged students to be active in their own academic success. Bill Hollister was the only teacher to do this.

Based on the second and third domains Bill Hollister is implementing formative assessments at a deep level. There is an incongruity between Bill Hollister’s understanding of formative assessment and his implementation of formative assessment. The low level definition in the first domain may indicate that Bill Hollister thinks of
formative assessment only as a task or simple check for understanding instead of as a process of understanding student thinking. Bill Hollister’s may perceive the continued questioning of students’ thinking and his modification of instruction in response to students’ thinking as a form of instruction completely separate from the formative assessment. Because of the mismatch between Bill Hollister’s understanding and implementation of formative assessment, I would characterize Bill Hollister’s implementation as being toward the middle of the deep implementation level.

**Developing Implementation**

There were four teachers, Trish Tisby, Pat Paulson, Chris Carter, and Julian Keller who exhibited Developing Implementation. These four teachers had indicators in the first two domains (Intentionality and Inquiry) that were high or positive but in the third domain, Modification of Instruction, there were no positive indicators. This group of teachers did not modify their instruction based on student data or take steps to promote active learning on the part of students.

**Intentionality of Integration**

In this first domain these four teachers each had some elements that indicated a level of intentionality that was more than trivial. This is not to say that the four teachers had matching profiles in this domain, but rather some indicators in this domain indicated that each of the teachers had some understanding above the lowest level.

**Definition and Example:**

Julian Keller gave the strongest definition of all of the teachers by including in her definition the idea that when a student doesn’t get something the teacher not only has to
reteach, but actually modify instruction. Julian Keller also gave an example that included teaching differently, but did not include the concept of understanding student thinking.

Pat Paulson gave a definition that included student understanding and re-teaching but the example given did not indicate modification in instruction or understanding student thinking. Trish Tisby’s definition was more basic with just checking to see if a student got it, but the accompanying example indicated a modification instruction although not understanding student thinking. Chris Carter also gave the basic definition of checking for if the student gets it, and the accompanying example only indicated re-teaching the content.

**Planning:**

In terms of including formative assessments as part of lesson planning Julian Keller usually did so while Trish Tisby and Pat Paulson sometimes did so. Chris Carter indicated that formative assessments were not included in her planning process.

**Barriers:**

Three of the four teachers in the Developing Implementation level indicated that time was a barrier to implementing formative assessments. Listing time as a barrier may be an indication that the teachers see formative assessment as an add-on to instruction, instead of as a method of instruction. It takes more time to do the formative assessments in addition to the instruction that they are already doing.

Pat Paulson indicated that the absence of a resource teacher was a barrier, because in the past the presence of a resource teacher allowed her to pull out students for small group and one on one work. Seeing the absence of a resource teacher as a barrier is
consistent with Pat Paulson’s definition of formative assessment, which included re-teaching as the response to when a student doesn’t understand.

**Inquiry and Interaction**

In this second domain all four Developing Implementation teachers used open tasks to begin the formative assessment process. Chris Carter asked “how many days old are you?”, Pat Paulson asked “how many cars do we need for a 3rd grade field trip,” and Julian Keller asked the students to distribute school supplies to all of the students at school. In each of these three cases, the students were coming up with multiple approaches to the questions. As each of these three teachers walked around and looked at student work, they ended up directing all of the students toward a single process, so by the time the whole group discussion happened, all of the students were doing their problems the same way.

As the teachers interacted with the students individually and in small groups, they did not explore the students’ thinking or have students explain their reasoning. Instead the teachers were very procedural, and moved students into following a specific set of steps to solve the problems. So, while their tasks were open, their questioning was procedural.

Trish Tisby used closed tasks and procedural questioning during her first observation. During Trish Tisby’s second observed class period the class worked on a single open task the whole class period. In that problem students, working in groups, were asked to redesign the classroom, choosing furniture and supplies for the new design. Because I did not see how Trish Tisby processed the lesson with the students, I did not witness whether the questioning of students was procedural or conceptual around this
task. Trish Tisby’s rating for the Task element is based on the second observation and
the rating for the Questioning element is based on the first observation.

All four teachers had above average times in assessment mode. Pat Paulson spent
and average of 37 minutes of class in assessment mode, which was the largest amount of
time for any of the teachers. Julian Keller followed closely with 35 minutes, Trish Tisby
with 30 minutes and Chris Carter with 23 minutes. All of these teachers were in
assessment mode for longer than the teacher average of 19.8 minutes.

These four teachers had very similar indicators in the second domain, but some
variance in the first domain. Julian Keller’s stronger definition and consistent planning
would put her at the top of this group, while Pat Paulson would be second based on those
same indicators. Trish Tisby would come next based on her recognition of needing to
teach differently in the example she gave. Chris Carter would be ranked lowest in the
Developing Implementation level based on the low level definition, example, and amount
of time in assessment mode.

The remaining teachers are going to be in the Superficial Implementation level,
and for a helpful visual representation, the teachers data can be seen by level in Table 9:
Depth of Implementation Levels on the following page.
<table>
<thead>
<tr>
<th>Name</th>
<th>Depth of Implementation</th>
<th>Intentionality of Integration</th>
<th>Inquiry and Interaction</th>
<th>Modification of Instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Definition</td>
<td>Examples</td>
<td>Planning</td>
</tr>
<tr>
<td>Bill Hollister</td>
<td>Deep</td>
<td>1- does student get it?</td>
<td>3 - have student work the problem, to figure out their thinkin, reteach using different strategies</td>
<td>1 - no</td>
</tr>
<tr>
<td>Julian Keller</td>
<td>Developing</td>
<td>3. student plus modify instruction</td>
<td>2 - teach differently</td>
<td>3 - usually</td>
</tr>
<tr>
<td>Pat Paulson</td>
<td>Developing</td>
<td>2 - student plus reteach</td>
<td>1 - reteach</td>
<td>2-sometimes</td>
</tr>
<tr>
<td>Trish Tilby</td>
<td>Developing</td>
<td>1- does student get it?</td>
<td>2 - teach differently</td>
<td>2-sometimes</td>
</tr>
<tr>
<td>Chris Carter</td>
<td>Developing</td>
<td>1- does student get it?</td>
<td>1 - reteach</td>
<td>1 - no</td>
</tr>
<tr>
<td>Natalie Night</td>
<td>Superficial</td>
<td>2 - student plus reteach</td>
<td>1 - reteach</td>
<td>1 - no</td>
</tr>
<tr>
<td>Howard Stenzel</td>
<td>Superficial</td>
<td>1- does student get it?</td>
<td>3. understand student thinking and give hints of prior learning</td>
<td>1 - no</td>
</tr>
<tr>
<td>McKenzie Orr</td>
<td>Superficial</td>
<td>1- does student get it?</td>
<td>1 - reteach</td>
<td>1 - no</td>
</tr>
<tr>
<td>Micah Norris</td>
<td>Superficial</td>
<td>1- does student get it?</td>
<td>1 - reteach</td>
<td>1 - no</td>
</tr>
<tr>
<td>Cecilia Reyes</td>
<td>Superficial</td>
<td>1- does student get it?</td>
<td>1 - reteach</td>
<td>1 - no</td>
</tr>
<tr>
<td>Julie Jones</td>
<td>Superficial</td>
<td>1- does student get it?</td>
<td>1 - reteach</td>
<td>3 - usually</td>
</tr>
<tr>
<td>Marie Montee</td>
<td>Superficial</td>
<td>1- does student get it?</td>
<td>1 - reteach</td>
<td>1 - no</td>
</tr>
<tr>
<td>Gwen Stacy</td>
<td>Superficial</td>
<td>1- does student get it?</td>
<td>1 - reteach</td>
<td>2-sometimes</td>
</tr>
</tbody>
</table>
Superficial Implementation

Eight of the teachers fell into this category of Implementation. In the Superficial level of implementation, in the Modification domain, teachers demonstrated no modifications of instruction or promotion of active learning by students as part of a formative assessment process. In the Inquiry and Interaction domain, these teachers only used closed tasks and procedural questioning as part of their formative assessment processes. There was some variety in the Intentionality domain.

Intentionality of Integration

For the eight teachers in the Superficial Implementation level, Intentionality of Integration is the only domain where teachers registered anything above the lowest indicators. The self reported indicators were generally low level and the time in assessment mode was generally below average. If a teacher indicated a mid or high-level response on any of the self reported elements, it was not reflected in the actual observed teaching.

Definition and Example:

Seven of the eight teachers in this category gave definitions of formative assessment that simply sought to determine whether a student gets a concept or not. One teacher, Natalie Night, gave a definition that included the need to reteach. When asked to give examples of what happens when students “don’t get it,” seven of the eight teachers gave examples where they re-taught the material without seeking student thinking or modifying the teaching. One teacher, Howard Stenzel, gave an example that included understanding student thinking, saying, “I have the student walk me through their
process, and catch their error. If they don't know where to start, I give them hints from prior learning.”

Planning:

Six of the eight teachers said they do not plan formative assessments in their lesson plans. Gwen Stacy responded to the question, “do you plan formative assessments in your lesson plans?” with “sometimes” while Julie Jones responded with “usually.”

Barriers:

When asked if there were barriers that hindered their use of formative assessment, two teachers, Julie Jones and Natalie Night, said there were none. Four of the teachers listed time as the barrier to effective use of formative assessments and peer pressure and the absence of a resource teacher were each mentioned once as the barrier to implementation.

Time in Assessment Mode:

When looking at the amount of time the teachers in the Superficial Implementation level spend in assessment mode, seven of the eight teachers are below the average of 19.8 minutes. Natalie Night was the only teacher in this level above the average time with 23 minutes of class time being spent in assessment mode. Gwen Stacy, Marie Montee and Julie Jones each had assessment times in the single digits with 8, 9 and 9 minutes respectively.

All of the low level indicators, along with the small amounts of time in assessment mode characterized Superficial Implementation of formative assessment.
Comparisons and Associations Between Depth Of Implementation Levels

Above I noted that the data from the teachers fell into stratification that had natural breaks. One division occurs between the Superficial Implementation and the Developing and Deep Implementations. This division is manifested in the tasks used to begin formative assessment. Teachers with Superficial Implementation did not use any tasks that allowed for students to think outside of a prescribed process. The teachers in Developing and Deep Implementation used open tasks, which allowed students to think through the math at hand in a way that made sense to them. The big distinction between Developing and Deep Implementation was whether the teacher capitalized on the students’ thinking through the use of conceptual questioning.

Bill Hollister, the teacher with Deep Implementation, solicited student thinking, and then capitalized upon the student thinking solicited through the open tasks and conceptual questioning by modifying his instruction and promoting an active learning environment for the students.

There is some association between the amount of time in assessment mode and the level of Depth of Implementation. (See Table 9: Depth of Implementation Levels). The single exception to this association is Bill Hollister, the teacher with Deep Implementation. While Bill Hollister had the deepest implementation, he only had the fourth longest time in assessment mode. Neither the literature nor the data of this study contain specific cutoffs regarding the average amount of time a teacher should be in assessment mode. The data does seem to indicate, however, that more time in assessment mode tends to yield deeper implementation.
While there doesn’t seem to be much association between a teacher’s definition and their level of implementation, there does seem to be some alignment between a teacher’s explanation of what they do when a student doesn’t understand and their level of implementation. This distinction highlights one of the disagreements found in the literature: whether formative assessment is a task or a process.

All of the teachers, in their examples, understood that when a student did not understand a concept the teacher would need to reteach in some capacity, but only three teachers included a component of re-teaching in their definitions. This would indicate that most of the teachers only think of the initial inquiry or task as the formative assessment, and their response to the data gathered from the student is something other than or outside of the formative assessment.

The data also indicates that even if a teacher has a low level definition of formative assessment, they can have deep implementation of formative assessment, as is the case with Bill Hollister.

*Teachers’ Self Ratings*

In the interviews I asked the teachers to rate themselves on their perceived effectiveness of their use of formative assessments. I had the ratings before I did any of the observations, and as I did the observations, I was increasingly surprised at the gap between the teachers’ self-ratings and the data that I was collecting. The eight teachers who demonstrated Superficial Implementation all gave themselves 6s and 7s on a scale of 1 to 10. The teachers’ self-ratings and the levels of implementation are not aligned, as displayed in Table 10: Teachers’ Self-Ratings below.
Table 10  
*Teachers' Self-Ratings*

<table>
<thead>
<tr>
<th>Name</th>
<th>Depth of Implementation</th>
<th>Teachers' Self-Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bill Hollister</td>
<td>Deep</td>
<td>6</td>
</tr>
<tr>
<td>Julian Keller</td>
<td>Developing</td>
<td>7</td>
</tr>
<tr>
<td>Pat Paulson</td>
<td>Developing</td>
<td>9</td>
</tr>
<tr>
<td>Trish Tisby</td>
<td>Developing</td>
<td>6</td>
</tr>
<tr>
<td>Chris Carter</td>
<td>Developing</td>
<td>3</td>
</tr>
<tr>
<td>Natalie Night</td>
<td>Superficial</td>
<td>7</td>
</tr>
<tr>
<td>Howard Stenzel</td>
<td>Superficial</td>
<td>7</td>
</tr>
<tr>
<td>McKenzie Orr</td>
<td>Superficial</td>
<td>6</td>
</tr>
<tr>
<td>Micah Norris</td>
<td>Superficial</td>
<td>7</td>
</tr>
<tr>
<td>Cecilia Reyes</td>
<td>Superficial</td>
<td>7</td>
</tr>
<tr>
<td>Julie Jones</td>
<td>Superficial</td>
<td>6</td>
</tr>
<tr>
<td>Marie Montee</td>
<td>Superficial</td>
<td>7</td>
</tr>
<tr>
<td>Gwen Stacy</td>
<td>Superficial</td>
<td>6</td>
</tr>
</tbody>
</table>

Further, when teachers were asked what they believed they could do to improve, teachers only gave two responses. The teachers said they would be more conscientious and deliberate in their use of formative assessment or that they would use formative assessments with more regularity. It is important to note that when teachers were asked to list things that would improve their scores, none of the teachers listed any of the indicators that would move them from Superficial Implementation to Developing Implementation or from Developing Implementation to Deep Implementation. Namely, none of the teachers mentioned using more open tasks, conceptual questions, understanding student thinking, modifying their instruction, or leading students to be more active in their own academic success as ways of improving their use of formative assessment.

After spending a lot of time with the data it occurred to me that the gap between the teachers’ self-ratings and their level of Depth of Implementation may be a function of
the definitions being used. I used the definition developed from the literature as a starting point for the Depth of Implementation Framework while teachers would naturally give themselves a rating based on their own definitions of formative assessment.

A reexamination of the Depth of Implementation Framework using each teacher’s own definition of formative assessment would probably yield different results.

**Depth of Implementation Based on Teachers’ Definitions**

During their interviews teachers gave a definition of formative assessment. Teachers then gave themselves a ranking from 1 to 10 on their own perceived effectiveness of the use of formative assessments. Teachers were also asked what they could do to improve. Ten of the teachers gave definitions that defined formative assessment as checking to see if a student can do the math. Two more teachers added a component of re-teaching, while one teacher included modifying instruction as part of the definition.

In response to the question, “what would you do to improve,” the teachers consistently gave two answers. First, that they would be more conscientious about using formative assessment, and second that they would use them more often. These two answers in response to improvement mean that teachers believe that planning for the use of formative assessments is part of using them effectively, and that there is a time component in terms of how often or how much time a teacher spends in assessment mode that makes a positive difference in the use of formative assessment.

The Depth of Implementation Framework has three levels while the teachers’ self-ratings were on a scale of 1 to 10. Because of this there is not a direct correlation between the two scales. For the purpose of re-examining Depth of Implementation using
teachers’ definitions, I will create a loose correlation. Superficial Implementation can be thought of as falling from 1 to 3. Developing Implementation would fall from 4 to 6, and Deep Implementation would start at 7 and go up to 10.

Looking at the teachers’ definitions of formative assessment, we see that ten teachers gave a definition that only requires the teacher, in some way, checks to see if the students understood the math. This definition does not require that the teacher solicit student understanding or modify instruct, but simply to check for understanding. Because each of these ten teachers used checks for understanding in their lessons, they would meet all of the requirements of their definition of formative assessment, and as such, would fall into the deep level of implementation (7 to 10 range).

The teachers identified more purposeful planning and more time in assessment mode as the ways to improve, so to determine where within the deep level of implementation these ten teachers would fall, we would look at whether they planned their assessments and how much time the spent in assessment mode, giving points for higher levels of planning and longer times in assessment mode.

For the three teachers who added of a component of instruction to their definitions, we have to ask if they met all of the elements of their definition. Two of the teachers Pat Paulson and Natalie Night, gave definitions that included an element of re-teaching. Both of those teachers retaught material as part of their formative assessment process, so they would be considered Deep Implementers and receive at least a 7.

Julian Keller’s definition of formative assessment included modifying instruction, but she did not modify her instruction during the classroom observations. This would mean Julian Keller only met half of the elements in her definition, and that
would put Julian Keller in the developing level at best, with a score between 4 and 6.

Julian Keller usually puts formative assessments in her lesson plans, and she spent 35 minutes in assessment mode, so both of these factors would raise her score within the developing level, giving her a 6. Table 11: Depth of Implementation Using Teachers’ Definitions has the teachers’ re-evaluated levels based on their own definitions.

<table>
<thead>
<tr>
<th>Name</th>
<th>Depth of Implementation</th>
<th>Definition</th>
<th>Planning</th>
<th>Time</th>
<th>Reteach</th>
<th>Modify Inst.</th>
<th>Rating</th>
<th>Level Based on Own Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bill Hollister</td>
<td>Deep</td>
<td>1- does student get it?</td>
<td>1- no</td>
<td>25</td>
<td></td>
<td></td>
<td>6</td>
<td>Deep - 8</td>
</tr>
<tr>
<td>Julian Keller</td>
<td>Developing</td>
<td>3. student plus modify instruction</td>
<td>3 - usually</td>
<td>35</td>
<td>Yes</td>
<td>No</td>
<td>7</td>
<td>Developing - 6</td>
</tr>
<tr>
<td>Pat Paulson</td>
<td>Developing</td>
<td>2 - student plus reteach</td>
<td>2-sometimes</td>
<td>37</td>
<td>Yes</td>
<td></td>
<td>9</td>
<td>Deep - 9</td>
</tr>
<tr>
<td>Trish Tilby</td>
<td>Developing</td>
<td>1- does student get it?</td>
<td>2-sometimes</td>
<td>30</td>
<td></td>
<td></td>
<td>6</td>
<td>Deep - 8</td>
</tr>
<tr>
<td>Chris Carter</td>
<td>Developing</td>
<td>1- does student get it?</td>
<td>1- no</td>
<td>23</td>
<td></td>
<td></td>
<td>3</td>
<td>Deep - 7</td>
</tr>
<tr>
<td>Natalie Night</td>
<td>Superficial</td>
<td>2 - student plus reteach</td>
<td>1- no</td>
<td>23</td>
<td>Yes</td>
<td></td>
<td>7</td>
<td>Deep - 7</td>
</tr>
<tr>
<td>Howard Stenzel</td>
<td>Superficial</td>
<td>1- does student get it?</td>
<td>1- no</td>
<td>15</td>
<td></td>
<td></td>
<td>7</td>
<td>Deep - 7</td>
</tr>
<tr>
<td>McKenzie Orr</td>
<td>Superficial</td>
<td>1- does student get it?</td>
<td>1- no</td>
<td>18</td>
<td></td>
<td></td>
<td>6</td>
<td>Deep - 7</td>
</tr>
<tr>
<td>Micah Norris</td>
<td>Superficial</td>
<td>1- does student get it?</td>
<td>1- no</td>
<td>15</td>
<td></td>
<td></td>
<td>7</td>
<td>Deep - 7</td>
</tr>
<tr>
<td>Cecilia Reyes</td>
<td>Superficial</td>
<td>1- does student get it?</td>
<td>1- no</td>
<td>10</td>
<td></td>
<td></td>
<td>7</td>
<td>Deep - 7</td>
</tr>
<tr>
<td>Julie Jones</td>
<td>Superficial</td>
<td>1- does student get it?</td>
<td>3- usually</td>
<td>9</td>
<td></td>
<td></td>
<td>6</td>
<td>Deep - 8</td>
</tr>
<tr>
<td>Marie Montee</td>
<td>Superficial</td>
<td>1- does student get it?</td>
<td>1- no</td>
<td>9</td>
<td></td>
<td></td>
<td>7</td>
<td>Deep - 7</td>
</tr>
<tr>
<td>Gwen Stacy</td>
<td>Superficial</td>
<td>1- does student get it?</td>
<td>2-sometimes</td>
<td>8</td>
<td></td>
<td></td>
<td>6</td>
<td>Deep - 7</td>
</tr>
</tbody>
</table>

When teachers’ use of formative assessment is evaluated based on the individual teacher’s definition of formative assessment, their scores are very close to their self-rating scores. Only one teacher, Julian Keller, received a lower score in this process than their self-evaluation score, but she also had the most complex definition of the teachers.
Chapter Summary

Implementing formative assessments in the classroom involves several actions on the part of teachers. The teachers have to solicit information from students, interpret that information, and respond to the information, all with the goal of improving student achievement.

The Depth of Implementation Framework is a theoretical framework that allows us to understand levels of implementation in a more complex way. The teacher who displayed deep implementation solicited information from students through open tasks that allowed students to approach problems in ways that made sense to them. The teacher allowed students to explain their thinking around the problem, and when there were misconceptions, the teacher empowered students to assist each other to help build understanding among all students, in a flexible learning environment. The teacher used a flexible and open process from start to finish and this allowed students to be active in their own and each other’s academic success. This teacher treated formative assessment as a mode of instruction, even though his own definition was a low level definition.

The four teachers who demonstrated developing implementation usually began their approach to students with an open task, but became more procedural through the questioning process. These teachers used formative assessments to gather information from students, but then steered students toward a specific process of math. They created the opportunity for students to approach problem in ways that made sense to the students, but did not use the students’ data to make the follow up decisions regarding modification of instruction.
The eight teachers who demonstrated superficial implementation treated their use of formative assessment as ancillary to instruction instead of as a mode of instruction. These teachers were very process driven, and the formative assessment was used as an occasional check to see if students could follow the process, and if not, the teaching of the process was simply repeated. These teachers also had the lowest amount of time in assessment mode in their class periods.

While most of the teachers performed in the lowest level of the Depth of Implementation Framework, as a group their self-ratings were much higher. This difference in scores can be attributed, at least in part, to how formative assessment is defined. When the Depth of Implementation analysis is repeated using teachers’ definitions, the teachers’ resulting depth of implementation levels are much more closely aligned to the self-evaluation scores.
CHAPTER 6: FACTORS IMPACTING USE OF FORMATIVE ASSESSMENT

Teachers never work in a vacuum. Their instruction is always influenced by a multitude of factors. Teachers will be aware of many of those factors, such as, whom they are teaching, what they are teaching, their own content and pedagogical knowledge to name a few. In the context of use of formative assessment, I was able to identify several factors that impacted teachers’ depth of implementation.

**Teacher Identified Barriers**

In the interviews with teachers I asked them to identify any barriers to their effective use of formative assessments. Eight teachers identified time as a barrier. Specifically, these teachers indicated that using formative assessments adds time to the lesson and it becomes more difficult to cover all of the material that has to be covered over the course of the school year.

Teachers’ description of time as a barrier gives insight into teachers’ mindsets with regard to formative assessment. These teachers viewed formative assessment as an add-on to instruction instead of instruction itself. If teachers have the mindset that formative assessment is not a form of instruction in and of itself, but just an add on, it may serve, as teachers indicated, to limit the amount of time that a teacher uses formative assessment. It may also serve to limit a teacher’s willingness to deepen their implementation of formative assessment.

Two of the teachers identified peer pressure between students as a barrier. They felt that the students would not always be honest about whether they understood something in front of their peers. If a teacher believes that they won’t get good information from a formative assessment, the teacher may be less likely to use it. Marie
Montee mentioned peer pressure and she only spent 9 minutes in assessment mode and had superficial implementation. However Bill Hollister also mentioned peer pressure as a barrier, and he spent 25 minutes in assessment mode with deep implementation.

Gwen Stacy and Pat Paulson both identified the absence of team teacher as a barrier. In both cases the teachers had worked with a team teacher in the previous year, which had allowed a lot more freedom to pull students from the whole class and work with them one on one or in small groups. The absence of a team teacher may have lowered the time in assessment mode for these two teachers. Gwen Stacy had the lowest time in assessment mode at 8 minutes, with superficial implementation. Pat Paulson, however, had the highest time in assessment mode at 37 minutes, at a level of developing implementation.

These were the only teacher-identified barriers to the effective use of formative assessments, but the data indicates that there may be other systemic and individual barriers.

**District Barriers**

Teachers were asked if they were getting enough district support for their effective use of formative assessment and all of the teachers responded in the affirmative. In response to this question Micha Norris said, “Yes, we get observed by the principal and the assistant superintendent. They give us feedback about our instruction and always support our using formative assessments.” (Micha Norris Interview, August 21, 2014). Cecilia Reyes responded saying, “yes, the principal is very supportive. When we have professional development in staff meetings we usually talk about formative assessments.” (Cecilia Reyes Interview, August 21, 2014) While all of the teachers saw the district’s
efforts as supportive, I believe the data indicates that these efforts are actually a barrier to improvement in the use of formative assessment.

There was some variance in the definitions of formative assessments between teachers, but the teachers with the low-level definitions were more likely to be implementing at the superficial level (7 of 10 teachers with low-level definitions). The correlation between low-level definitions and superficial implementation raises two issues regarding how the district supports formative assessment in the classroom: that the district has not conveyed a specific definition of formative assessment, and that it allows low level definitions of formative assessment to exist within the district.

The district, through the principal and assistant superintendent, has encouraged teachers to use formative assessments in the classroom. The district has provided professional development around formative assessment, however the principal indicated that the district has not conveyed a specific definition of formative assessment.

The data tell us that teachers seem to base their implementation of formative assessment on their definitions of formative assessment. When asked what they could do to improve their implementation, teachers indicated that they could plan their formative assessments more consciously or use formative assessments more often. It is important to note that teachers did not point to any elements outside of their definitions as a source of improvement. There were no teachers with low-level definitions who said making sure they understood student thinking was how they could improve their use of formative assessment. The low level definitions are a barrier, and because the district does not address this in its professional development, the lack of a shared definition becomes a district barrier problem.
The second issue regarding how the district supports formative assessment is that it allows the low level definitions to exist. It is conceivable that even though there is a lack of a shared definition of formative assessment within a district, the teachers could each have their own high level definition which included elements of understanding student thinking, modifying instruction and empowering students in their own learning. The lack of a shared definition, in and of itself does not mean that teachers necessarily have low-level definitions. In this case, however, the majority of teachers do have low-level definitions. The norm for the teachers in the study is a low level definition (10 of 13), and the professional development along with feedback from administrators has not changed teachers’ definitions. The district’s efforts to support formative assessment have allowed these definitions to persist, and as such, reinforce the barrier to improvement that the low level definitions present.

Other Teacher Barriers

While the persistence of low-level definitions is a district barrier, the low-level definitions, in another way, are also a barrier for individual teachers.

Pat Paulson gave a self-rating of 9, and when asked what she could do to improve she said, “I’m not sure. I feel we can always improve. Perhaps add more types of formative assessments.” Pat Paulson’s response was not an expression of ignorance of formative assessment, but rather an acknowledgement of her belief that she was already performing at a very high level, and there was not much room for improvement. It is this sentiment from Pat Paulson, and to a lesser degree with the other teachers, that becomes the barrier.
If a teacher believes that she has a strong understanding of a particular methodology, and that she is implementing that methodology well, she has very little incentive to do research or learn more about that methodology. All of the teachers expressed their definitions unhesitatingly and with confidence. Each teacher gave an above average self-rating, with the exception of Chris Carter who gave herself a 3. When Chris Carter was asked what she could do to improve her use of formative assessment, she said, “I could do them more often, because I don’t always do them.”

When asked about how to improve, none of the teachers questioned their own understanding of formative assessment. No one mentioned “doing research on the topic” or “deepening my understanding of the formative assessment methodology” as ways to improve their own implementation. Each teacher believed that their understanding of formative assessments was sufficient, and this belief is a barrier to deepening their implementation of formative assessment. None of the teachers was inclined to question their own definition as a part of improvement, and while the teachers acknowledged that they could go to their principal for assistance or coaching the principal indicated that none of the teachers had done so with regard to formative assessment.

**Chapter Summary**

Teachers identified time, peer pressure, and the absence of a resource teacher as barriers to implementing formative assessments in their classrooms. Time was the most identified barrier, and that fact may indicate that teachers see formative assessments as something that takes additional time to add to instruction instead of the instruction itself. This mindset would be a barrier to improving the use of formative assessments.
The lack of a high level definition for formative assessment also acts as a barrier to improvement. Each teacher can use whatever definition they want, and there are no mechanisms in place to help teachers reexamine the definition as part of an improvement cycle. When teachers identify ways they can improve, they do not point to practices outside of the definitions that they already hold.

The implications of this study for improving teachers’ use of formative assessment, and other methodologies that have developed organically in schools and aren’t tightly defined, are discussed in the next chapter.
CHAPTER 7: IMPLICATIONS FOR PRACTICE, POLICY, AND RESEARCH

This is a study of the understanding and use of formative assessments of thirteen teachers in a small suburban district. Teachers were interviewed and observed over a three-month period. This was a small-scale study, and the design and size of the study necessarily create limitations. While the study itself has value, and creates implications for further research, the size and qualitative design mean that the findings are not generalizable to other districts or the larger field of teaching.

The teachers who participated in this study all willingly volunteered to be a part of the study. Their willingness to volunteer likely points to their comfort level with their use of formative assessment. Each of the volunteer teachers expressed an understanding of formative assessment, so the teachers with doubts about formative assessment probably did not volunteer, which would skew the data as compared to a completely random sample of teachers.

Some of the data was collected through observation, and the act of observation probably had some influence on what teachers did. The teachers knew that I was looking for formative assessments, and this knowledge may have nudged teachers to use formative assessments more than they normally would have.

Studying teachers understanding and use of formative assessment across a variety of districts over a longer period would give a more generalizable result, as would a wide scale study of teachers’ use and understanding of a variety of teaching methodologies that suffer from a lack of a standard shared definition (such as differentiated instruction, instructional scaffolding, or cooperative learning).
I used thematic analysis to develop a Depth of Implementation Framework based on multiple data points from teachers. That data was gathered through both individual interviews and multiple observations. I used several data points to ensure validity, but validity may have been heightened by gathering additional data through the same avenues over a longer period of time, plus the use of additional avenues such as survey instruments or focus groups. I also only had one teacher fall into the deep level of implementation. Because of this, I did not do any analysis between teachers within the deep implementation level.

For this study I did not look at any student academic outcomes in relationship to teachers’ use of formative assessments. This study was, in a sense, a formative assessment of teachers’ use of formative assessment. Ultimately the goal of using formative assessments is to improve student achievement, so the addition of student achievement data in further research would add the dimension of determining if, and to what degree there is a link between teachers’ depth of implementation and changes in student achievement.

**Implications for Practice**

While this study is of a limited scope, it produced some findings that may be used by educators to inform practice.

When new scripted programs are implemented in school districts, there is usually an “instruction manual” that comes with it. The creators of the new program will often strongly define how the new program should be implemented, and what indicators should be used to measure if teachers have implemented the new initiative with fidelity (Baker, Liberman, & Kuehnel, 1986; Lapan, Gysbers, & Sun, 1997). This study looked at
formative assessment, an educational methodology that does not have a specific author or script and is not strongly defined. The absence of the “instruction manual” for formative assessment helps to make implementation inconsistent. School districts should be able to improve teachers’ use of formative assessment by being more scripted in its implementation.

**Do Not Assume Shared Understanding**

The use of formative assessments in the classroom is not a new initiative, and each of the teachers in this study had an understanding of formative assessment from which they were operating. All of the teachers also reported that the district encouraged and supported their use of formative assessments, and provided professional development about formative assessment, but there was no specific shared definition of formative assessment, nor a shared process for its use.

The data from this study points to the idea that a teacher’s implementation of formative assessment is based, in some part on the teacher’s definition of formative assessment. The implication for educators who are working to improve instructional practice with formative assessments is that they cannot assume that teachers have a shared understanding of formative assessment.

Districts that are encouraging the use of formative assessment should define formative assessment and create a framework for implementation so all teachers have a shared understanding of what it is and how it should be used. A framework for implementation helps teachers understand what elements necessarily need to be present for effective formative assessments, and give some guidance on how each of the elements should be used.
While this study only looked at formative assessment, educators can also apply this implication to other established teaching methodologies such as differentiated instruction or instructional scaffolding, and always go back and establish with teachers a shared understanding of the definitions and implementation strategies of methodologies.

**Use Multiple Data Sources to Evaluate Teachers’ Use of Formative Assessment**

Looking at teachers’ use of formative assessment through the Depth of Implementation Framework, eight of the teachers were categorized in the superficial implementation level, and only one teacher reached deep implementation. When those same teachers’ use of formative assessment was reviewed using their own definitions of formative assessment, they were rated much higher. Knowing how the teachers understood formative assessment made a difference in how their implementation of formative assessment could be evaluated. Educators should think about collecting data on teacher performance from multiple sources instead of just observing a teacher using formative assessments or any other established methodology.

In this study there was a loose association between time in assessment mode and depth of implementation. If educators collect data from several sources when evaluating implementation of methodologies, they can get a much better sense of how teachers are implementing those methodologies, which also informs how and what types of professional development might be used to improve practice.

**Differentiate Teacher Professional Development**

The levels of implementation of formative assessment in this study ranged from superficial to deep. Within the superficial and developing levels there were differences in how teachers understood formative assessment and how teachers were implementing
formative assessment. This diversity in the understanding and implementation of formative assessment imply that different teachers need different things to deepen their implementation. Two of the four teachers in the developing level indicated that they understood they should be modifying instruction, while the other two only expressed the need to reteach.

Differentiating teacher professional development and instructional supports will give teachers the support that they specifically need instead of using a one size fits all approach to professional development, which forces teachers to sit through training that may not be helpful to move that teacher to a deeper level of implementation. Currently in the district teachers are observed individually as part of the assessment process, but professional development is done with the whole school regardless of where teachers are in their understanding and implementing the various initiatives that are present in the district.

Address Time As a Perceived Barrier

Eight of the teachers in the study identified time as a barrier to their effective use of formative assessment. The specific issue as they identified it was that using formative assessments consistently meant that they were losing the time they need to teach the material. These teachers expressed this issue in such a way that it was clear that they saw formative assessment as an add-on to instruction instead of a method of instruction. Bill Hollister, the deep implementing teacher did not identify time as a barrier, but three of the four teachers at the developing implementation level did.

The implication here is that teachers who have a developing implementation may still see formative assessment as an add-on to teaching instead of as a way of teaching.
This mindset may persist across all implementation levels, and the result of this may be that even if teachers are good at formative assessment, they may choose not to do it very often because they see it as a trade off to instruction instead of instruction itself.

Addressing this mindset becomes critical if the district wants teachers to consistently use formative assessments in their classrooms.

As part of professional development, the district can do some things to address this mindset, including using Bill Hollister as an example of deep implementation where the formative assessments were used as instruction and not treated by the teacher as something that is added to instruction.

*Teachers As Self-Directed Learners*

Ideally, if a school or district uses a Depth of Implementation analysis on formative assessments and other existing methodologies that are not strongly defined, teachers will get in the habit of using the framework to self-assess with all of their methodologies. In this study, teachers did not look outside of their own definitions when thinking about improving their practice. Teachers can use the Depth of Implementation Framework to assist in self-assessments. The framework will help them to ask questions regarding what are the necessary elements of a given methodology, and how do you determine if each of those elements are being implemented deeply.

If teachers begin to ask those questions of their own practice, and push the district administrators to be more explicit with their documentation of instructional methodologies, then this research also holds implications for instructional policy.
Implications for Policy

It has been my experience in education that at the district level, school leaders do not have a strong sense of how policy affects the day-to-day workings of a district. Policy is often used to define and discourage bad behavior, and much less often used to define, encourage, and support desired behaviors. The Central School District is no exception.

The teachers in the Central School District understood that they should be using formative assessments. The district, however, did not have any policy statements/documents that helped teachers understand what formative assessment is and how to use it. It would also be helpful to teachers if the district include a description of the supports that a teacher could expect to access as they try to deepen their use of formative assessments. If a district includes the use of formative assessments in their educational policy statements, it will be helpful to teachers if those policies include a fully developed definition of formative assessment, and give some description of what deep implementation looks like so teachers know what they are supposed to be doing.

The district should also be mindful of crafting policies that not only define formative assessment, but that also describe where formative assessment fits in the gallery of teaching practices that are addressed in policy. In this study a number of teachers saw formative assessment as an add-on to their instruction instead of as part of the instruction itself. Often, district mandates on instructional practice are seen as disparate ideas that are in competition with each other for both time and resources. Teachers may see instructional mandates as “just one more thing we’re expected to do”
instead of a cohesive set of tools which allow teachers to maximize the effectiveness of their instructional practice.

Mapping out how formative assessment interacts with other instructional methodologies and mandates should help teachers incorporate formative assessments more effectively in their classroom instruction. The Central School district can use policy not to define how those methodologies work together to define effective instruction.

**Implications for Research**

Formative assessment is not a new methodology, but this study has highlighted the fact that there is not a shared understanding of formative assessment by the teachers who are purportedly using it on a daily basis. More research on the topics of teacher understanding and use of formative assessment can help pinpoint the disconnects that allow such a variance in understanding, and it can help teachers and districts begin to prevent or at least lessen the variance when asking teachers to use formative assessments in classrooms.

Using a Depth of Implementation Framework in continued research of formative assessments can help educators to better define when formative assessment is being used effectively. In this study, I only had one teacher who had deep implementation of formative assessment. The other teachers in the study fell into two levels that were easily defined. Further research using this framework can help to define the levels of implementation and differentiate formative assessment use within each level. With more teachers participating in the research, we may find that there should be four or five levels of implementation. The more finely the use of formative assessment is differentiated, the
more targeted the supports can be to help teachers deepen their use. Further research along these lines should give educators the tools to increase their effective use of formative assessment.

Using a Depth of Implementation Framework in other established teaching methodologies can also be helpful to the implementation of those methodologies. There are a number of methodologies in education that are not new programs. Many of these, such as differentiated instruction or instructional scaffolding suffer from the same issues as formative assessment. These older methodologies are often poorly defined, and while educators know they should be using them, they often do so superficially, and see the use of these methodologies as an add-on to instruction instead of as the instruction itself.

By using a Depth of Implementation Framework with a variety of methodologies, educators better understand how those methodologies are defined and implemented. By using a single framework across methodology research, it also makes the research much more accessible to educators who do not have to figure out a new framework or construct with each new piece of research. Having a single framework for a variety of methodologies would allow educators to use a consistent process of implementation and simplify the learning curve when trying to implement or improve the use of a methodology.

Doing more in-depth research that ties together teachers’ understanding and use of educational methodologies with a single Depth of Implementation Framework can give districts more and better information about how to implement, evaluate the use and support of implementation, and increase the depth of implementation of a variety of
instructional methodologies in a consistent manner. This should ultimately lead to better instruction.

**CONCLUSION**

The research data on the positive effect of formative assessment on student achievement has existed for nearly two decades, but in recent years that research has come under scrutiny for a number of reasons, including the absence of a consistent lexicon and the absence of a research validated framework for formative assessments. Most schools continue to ask teachers to use formative assessments, even in the absence of lexicon and framework. This ambiguity may in fact make it more difficult for teachers to effectively implement formative assessments in their classrooms.

This study examined teachers’ understanding and use of formative assessment. The absence of a shared understanding and definition of formative assessment did have an effect on how teachers implemented formative assessments in their classrooms. The understanding of formative assessment that each teacher happened to have became both the driver and limiter for how deeply a teacher implemented formative assessments. Teachers with higher-level definitions tended to implement formative assessments more deeply. Teachers who used formative assessments more often also had deeper levels of implementation.

A teacher’s definition of formative assessment also limited a teacher’s ability to envision improvement strategies for their implementation of formative assessment. To improve they suggest being more mindful with what they were already doing or doing what they were doing more often. The teachers did not look outside of the definition they already held to find improvement strategies.
If school administrators want to continue encouraging teachers to use formative assessments, they should begin with creating a common definition so every teacher has a shared understanding of what a formative assessment is. The administrators should also create a framework that looks at multiple data points around a teacher’s understanding and use of formative assessment so the process of improvement can be fine tuned and supports for teachers can be targeted to the specific areas where individual teachers need assistance. In a sense, administrators need a formative assessment for teachers use of formative assessment, or any other methodologies being implemented.

Continuing research in classrooms, on the actual use of formative assessment will serve to create a better framework for best practices, and remove much of the ambiguity around formative assessment that is currently a barrier to implementation.
APPENDICES
Appendix A
Participant Informed Consent Form

Study of Elementary Math Teachers’ Understanding
and
Use of Informal Formative Assessments

Participation is Voluntary:
It is your choice whether or not to participate in this research. If you choose to participate, you may change your mind and leave the study at any time. Refusal to participate or stopping your participation will involve no penalty or loss of benefits to which you are otherwise entitled.

Study Description:
The dual purpose of this study is, first, to understand the lived experiences of elementary math teachers use of formative assessments in their classrooms; and second, to view these experiences in concert with a framework that I developed and will modify based on what I learn through this research.

How Long Will I Take Part?
Your participation will involve two interviews of an hour each, and two classroom observations of a math lesson over the course of two to three months.

Expectations of Participants:
If you choose to be a part of this study, you will be interviewed on two occasions for about an hour each time, and between the interviews, on two occasions, I will observe you teaching math to students. The interviews will be recorded and the observations will not be recorded. I will take notes in both the interviews and the observations.

Risks:
If you choose to participate in this study, there are no major risks associated with the study.

Confidentiality:
Your confidentiality will be respected throughout the process. All participants and the district will have a pseudonym, and only those pseudonyms will appear on any written notes and in the dissertation itself. All of the information that I collect will be kept in a secure place at my home, or under password protection on my computer. I will take every precaution not to use any information that would identify any individual. At the conclusion of the research I will destroy all of the audio recordings of the interviews.

Contact Information:
If you have questions or concerns about this study, please contact the researcher, Steven Harris, at the following: uspharris@gmail.com, (816) 726 5186, or 1552 NE Woodland Shores Way, Lee’s Summit, MO 64086.
My Harvard faculty advisor is Dr. Eileen McGowan, and she can be reached at: eileen_mcgowan@gse.harvard.edu, 617.496.8836, or Gutman 450, Harvard Graduate School of Education, 6 Appian Way, Cambridge, MA 02138.

This research has been reviewed by the Committee on the Use of Human Subjects in Research at Harvard University. They can be reached at 617-496-2847, 1414 Massachusetts Avenue, Second Floor, Cambridge, MA 02138, or cuhs@fas.harvard.edu for any of the following:

- If your questions, concerns, or complaints are not being answered by the research team,
- If you cannot reach the research team,
- If you want to talk to someone besides the research team, or
- If you have questions about your rights as a research participant.

Statement of Consent:

I have read the information in this consent form. All my questions about the research have been answered to my satisfaction.

Your signature below indicates your permission to take part in this research. You will be provided with a copy of this consent form.

_________________________  _____________________
Printed name of participant

_________________________  _____________________
Signature of participant      Date

I have discussed with _______________________ the above procedures, explicitly pointing out the potential risks and discomforts. I have asked whether any questions remain and have answered those questions to the best of my ability.

_________________________  _____________________
Steven E. Harris      Date
Interview Questions

Introduction:

My name is Steve Harris and I am doctoral student at Harvard’s Graduate School of education. Thank you for taking part in my research study.

I would like to interview you to learn more about your classroom use of informal formative assessments. The interview will last about one.

You can skip questions that you do not want to answer or stop the interview at any time.

I will keep the data I collect confidential, and will not share your personal information with anyone.

Being in this study is voluntary. Please tell me if you do not want to participate.

Questions?

*Review consent form.

*Ask permission to record interview.

Biographical Information: (To get into the rhythm of the interview):

“Tell me about your background as an educator?”

How long in each education role?

Current job and responsibilities?

Teacher’s understanding of informal formative assessment (Research Question 1 - ask for specific examples in math throughout):

When you hear the word assessment, what comes to mind?

Do you make a distinction between formative and summative assessments? What is that distinction? (Use teacher’s language if they made the distinction)
(Ask follow up questions to get a clear picture of their understanding of the terminology)

During a lesson, how do you know if a student has mastered the idea or concept presented? (elicit examples of non-written formative assessments)

Are there any math specific (formative assessments) that you’re aware of?

**Teacher’s Use of Formative Assessments (Research Question 2 -ask for specific examples in math):**

What things do you do to check if students grasp a concept in a lesson? (What are the formative assessment techniques that you use in math class?)

Can you give me some examples of when you did this (checked for understanding) in class this last week?

What sorts of data do you get through your checking for understanding (use whatever language the teacher uses for informal formative assessments)?

Do you have any examples of when you checked for understanding and a student didn’t understand?

When the student didn’t understand, what did you do?

What do you believe is the effect that your checking for understanding has in your class?

Before you start a lesson, do you plan out your checks for understanding?

Do you use the same checks for understanding for every student or is there some differentiation?

How do you decide when and how to differentiate your checking for understanding?
Resources and Supports (Research Question 3):

What resources or supports do you receive to help you to use (informal formative assessments) in class?

What resources or supports would you like that you are not getting?

Barriers (Research Question 3):

Are there checks of understanding that you would like to use but don’t use in class?

What things that prevent you, from doing those checks for understanding (or are there things that get in the way your using your checks for understanding the way you want to?)

Other Thoughts:

Are there any other things you can share with me regarding your use of formative assessments?
Harvard University-Area
Committee on the Use of Human Subjects
1414 Massachusetts Avenue, 2nd Floor
Cambridge, MA 02138
Federal Wide Assurance FWA00004837

Notification of Exemption Determination

November 18, 2013

Steven Harris
seh916@mail.harvard.edu

Protocol Title: Teachers’ Understanding and Use of Informal Formative Assessments in the Mathematics Classroom
Protocol #: IRB13-2760
IRB Review Date: 11/18/2013
IRB Review Action: Exempt

Dear Steven Harris:

On 11/18/2013, after review of your submission, the Institutional Review Board (IRB) of the Harvard University-Area determined that the above-referenced protocol meets the criteria for exemption per the regulations found at 45 CFR 46.101(b)(1) and (2).

Additional review by the IRB is not required. However, any changes to the protocol that may alter this determination must be submitted for review via a modification (by selecting the create modification/CR activity in the ESTR system) to determine whether the research activity continues to meet the criteria for exemption.

The IRB made the following determination:
- Research Information Security Level: The research is classified, using Harvard’s Data Security Policy, as Level 2 Data.

If you have any questions, please contact me at 617-495-1775 or fennever@fas.harvard.edu

Sincerely,

Fanny Ennever
Senior IRB Administrator
# Observation Form

<table>
<thead>
<tr>
<th>Time</th>
<th>Formative Assessment</th>
</tr>
</thead>
</table>

Teacher ___________________________  Date _____

Appendix D
REFERENCES


*Instructional Science, 18*(2), 119-144.


VITA
STEVEN E. HARRIS

Education

2007 – Present  Doctorate of Education Candidate, Graduate School of Education, Harvard University

June 2008  MA, Education Administration and Management, Graduate School of Education, Harvard University

August 1999  MA, Mathematics Education, New York University

May 1992  J.D., Boalt Hall School of Law, University of California at Berkeley

May 1987  B.S., Mathematics and Computer Science, William Jewell College

Professional Experience

2010 – Present  Owner and Principal, Harris Consulting

2013-2014  Adjunct Lecturer, Kennedy School of Government, Harvard University

2009 – 2010  Assistant Superintendent of Human Resources, Kansas City, Missouri School District

2008 – 2009  Special Assistant to the Superintendent, Prince George’s County Public Schools

2005 – 2007  Assistant Professor, Department of Education, William Jewell College


2002 – 2004  Owner at Harris Law Firm, Springfield, Missouri

2000 – 2002  Executive Director, Open House Ministries, Homestead Florida

1997 – 2000  President/ Immigration Director, Rauschenbush Metro Ministries, New York, New York
1998 – 1999  
High School Math Teacher, Hunter College High School, New York, New York

1996 -1997  
Attorney, Spencer Fane Britt & Browne, Kansas City, Missouri

1994 – 1996  
Director of Diversity, Central Missouri State University, Warrensburg, Missouri

1993 – 1994  
Asst. General Counsel, Missouri Department of Elementary and Secondary Education, Jefferson City, Missouri

1992 – 1993  
Attorney, Shook Hardy and Bacon, Kansas City, Missouri

Publications


Presentations