STUDENT LEARNING OF MEASUREMENT AND SOUND: EXAMINING THE IMPACT OF TEACHER PROFESSIONAL DEVELOPMENT

by

Ting-Ho Huang

B. S., National Taiwan University, 1998

M. A., University of Pittsburgh, 2004

Submitted to the Graduate Faculty of

School of Education in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2009

UNIVERSITY OF PITTSBURGH

SCHOOL OF EDUCATION

This dissertation was presented

by

Ting-Ho Huang

It was defended on

September 8, 2009

and approved by

Jennifer L. Cartier, Assistant Professor, Department of Instruction and Learning

Jana M. Iverson, Associate Professor, Department of Psychology

Carl N. Johnson, Associate Professor, Department of Psychology in Education

Dissertation Advisor: Ellice A. Forman, Professor, Department of Instruction and Learning

**STUDENT LEARNING OF MEASUREMENT AND SOUND:
EXAMINING THE IMPACT OF TEACHER PROFESSIONAL DEVELOPMENT**

Ting-Ho Huang, PhD

University of Pittsburgh, 2009

The current trends in science education emphasize students' engagement in scientific inquiry and deemphasize memorization of factual knowledge. In response to these trends, reform-based curricula often characterize students engaging in science investigations by listing the detailed steps of scientific practices. Unfortunately, if curricula stress procedures more than conceptual connections, students' understanding of science may be distorted. To ensure the effectiveness of science education reform, this study was designed to bring attention to the differences between the procedures for carrying out investigations and the conceptual knowledge involved in doing the investigations.

This study explored teachers' use of a hands-on science curriculum and its impact on student learning using multiple assessments. The relationships between student learning of conceptual or procedural knowledge and teachers' talk about conceptual or procedural knowledge in their classrooms were investigated. Three teachers and their third-grade students participated in this study during the teaching of the measurement and sound units from a inquiry-based curriculum. One teacher had participated in an enhanced professional development workshop but the other two comparison teachers did not.

The results showed that students from the three classes displayed similar knowledge of measurement procedures but different conceptual understanding of measurement and sound properties. Results from classroom observations indicated that teachers enacted the curriculum differently by either stressing the procedures of the activities or making conceptual connections between the activities and content knowledge. One of the comparison teachers emphasized the step-by-step procedures for using measurement tools, whereas the other two teachers constantly made reference to related measurement concepts or to everyday experiences. The patterns between student learning and teachers' enactment suggested that higher percentages of conceptual talk in the classroom were associated with better performance on conceptual understanding.

The results of this study are consistent with prior research that showed that the implementation of hands-on science curricula could result in limited instruction in science concepts and reduced conceptual learning by their students. These findings suggest that it is important to help teachers focus on conceptual as well as procedural knowledge when using an inquiry-based curriculum to enhance the quality of science education reform.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I would like to express my earnest appreciation to my academic advisor, Dr. Ellice Forman, who has been very supportive and patient throughout the entire process. She challenged me to go deeper about the current results and discover further about new findings. Many times when I felt frustrated, her knowledge and enthusiasm about research has always inspired and encouraged me to look beyond small obstacles and focus on big pictures. What I have learned from her is invaluable in both research and life.

I am also very grateful for the research experiences I gained while I was involved in Dr. Jennifer Cartier's multiple research projects. It has been a great privilege to work with her. Her creative ideas and dedication in both research and teaching have significantly broadened my view of science education.

I would like to thank my committee members, Dr. Jana Iverson and Dr. Carl Johnson, for their willingness to be in my committee and their excellent advice which helped to strengthen my research findings and arguments.

Many thanks to my colleagues and members of our support and study group, Linda Messineo, Brian Wargo and Yukari Yamakawa. We shared our research ideas, preliminary coding schemes, non-publishable research results, laughs, anxiety and whining together during this journey. It was wonderful to be part of this support group.

Over these years, my father: Jang-kai Huang, mother: Ching-hua Huang, and sister: Chen-ho Huang, have continuously showed their unconditional love and belief in me. My dear friends who provided consultations, sincere advice, warm greetings, and even delicious food have filled my heart with warmth along this journey. Without their support and encouragement, I would not have imagined that I could accomplish this degree.

I am indebted to so many people that guided me, stood beside me, or helped me in different ways. I would like to dedicate this dissertation to them as a way to convey my sincere and heartfelt gratitude for their great contribution to my life.

1.0 INTRODUCTION

The prevalence of science and its application in technology has widely affected our lives and made it impossible to avoid using scientific information in our daily conversations. New discoveries and scientific results reported by media require our best judgments to verify their plausibility and to apply cautiously those research findings. It is especially important to prepare the next generations to be able to comprehend, verify and apply the conclusions of these research findings. Therefore, the goal of science education, as stressed in many reform documents, is to educate students to become scientifically literate. This means students need to be able to apply scientific concepts and methods to make informed decisions about scientific information in their everyday lives (National Research Council [NRC], 1996). This trend presents challenges for science educators at different levels to shift from previous emphasis on covering numerous amounts of scientific facts and theories to considering the practical application of scientific thinking outside of the classroom.

1.1 STATEMENT OF THE PROBLEM

Recent science education reforms emphasize scientific thinking skills more than the memorization of factual knowledge (American Association for the Advancement of Science [AAAS], 1993; NRC, 1996; Schauble & Glaser, 1990). Science should be learned as an active process that students engage in authentic scientific practices and conduct scientific investigations, such as making observations, describing phenomena, proposing questions, designing experiments, collecting data, finding data patterns, constructing conclusions and explanations, and communicating results with others (AAAS, 1993; NRC, 1996). Simply teaching scientific facts and theories is no longer considered practical or efficient if we want students to apply scientific knowledge or scientific thinking skills outside of school settings. Reformers suggested that students should be engaged in the activities of scientific inquiries and exposed to environments that encourage the practice of scientific thinking, the conduction of scientific activities and the understandings of the nature of science.

The emphasis on students' abilities to inquire about science described in the national standards also asks for significant changes to science curricula and needs for professional development and appropriate assessment tools (NRC, 1996). Reform-based science curricula feature hands-on science activities in which students create or observe first-hand phenomena or solve tasks by applying relevant science knowledge. These hands-on activities are intended to create meaningful contexts of scientific inquiries so that students could engage in activities of scientific thinking, such as exploring causal relations, explaining possible causes, and evaluating and revising explanations. Despite the rich opportunities for inquiry that a curriculum may create, they require teachers' careful implementation to assimilate hands-on activities to scientific inquiries (Dalton, Morocco, Tivnan, & Mead, 1997; Huber & Moore, 2001; Marx et al., 2004). However, these curricula often stress procedures in a step-by-step format for teachers to easily follow (Huber & Moore, 2001) and provide little curricular support for inquiry-based instruction (Dalton et al., 1997).

Teachers' adaptation to inquiry-based instruction could critically affect the accomplishment of educational reform (Anderson & Helms, 2001; von Secker & Lissitz, 1999). Considering the lack of inquiry support in science curricula, teachers are challenged with utilizing the opportunities of scientific inquiries and guiding students to make sense of hands-on activities. When there is little support for inquiry-based instruction, it is likely that some teachers may follow faithfully the exact procedures of crafting materials and manipulating equipments but do not spend time on scientific practices that followed (e.g., search data patterns to find an explanation). The kinds of activities students actually engage in (i.e., the enacted curriculum) might not share the same themes as those anticipated by the curriculum developers (i.e., the published curriculum) (Remillard, 2005). The mere involvement of making materials to create phenomena does not connect to any scientific thinking without proper teacher guidance and may result in students learning nothing about science. Therefore, it is important to unveil teachers' actual enactment of the curriculum and to clarify the impact of teachers' enactment on students' learning.

1.2 BACKGROUND

This study is part of a larger project aimed at promoting effective science teaching by implementing critical elements of inquiry-based instruction in elementary school science (Cartier, 2003). In this larger study, called Pittsburgh Public Hands-On Science Study (PPHOSS), Cartier worked with six urban public school teachers (in grades K-5) collaboratively for two and a half years around a hands-on science curriculum. During those professional development workshops, teachers participated in activities that simulated scientific inquiries and were involved in discussions of organizing their instructions with scientific inquiries. Teachers were also introduced to Big Ideas, Tools and Talk as critical elements of inquiry-based instruction. Each teacher selected a focused FOSS unit based on their teaching grades and identified Big Ideas (or key conceptual models) within and across various activities to help students build and develop basic science concepts. Teachers were also encouraged to use Tools (e.g., tables, graphs, illustrations or object models) to summarize, organize or represent data patterns and to explore possible relations among observed data. In addition, teachers engaged students in Talk to describe observations and explain thoughts so that teachers could monitor student understandings. Throughout the workshops, teachers discussed and presented possible implementation of Big Ideas, Tools and Talk (BITT) to enhance inquiry teaching of the focused modules.

The aim of this dissertation study was to evaluate the impact of teachers' enactment of the curriculum on students' learning, with or without BITT training. One teacher from PPHOSS was selected for this study (the target teacher) based on her teaching duty at the time of data collection and her willingness and readiness to participate. In order to compare how FOSS was actually implemented in elementary public schools, this study included two other teachers (comparison teachers) who did not learn about BITT but taught the same grade (third grade) as the PD teacher. These two comparison teachers were selected by their district supervisors as implementing the FOSS curriculum faithfully. The varieties of teachers' enactment of the curriculum may have significant impact on students' learning of science and will be explored in this study.

This study focused on two FOSS third-grade modules: Measurement and Physics of Sound modules. The Physics of Sound module was one of the discussion modules in the PPHOSS workshop and was thus selected as the target module in the current study. In order to take into account students' differences prior to the intervention of PPHOSS, the Measurement module, which was taught at the same grade before the Physics of Sound module, served as comparison index about students' learning in the three classrooms. Additionally, the findings of students' learning of Measurement and Sound will build onto the literature of these two content areas.

1.3 THE STUDY

1.3.1 Research questions

The use of activities-based curriculum indicates the direction that science should be learned through participating in the activities of scientific inquiries and students should understand the nature of science by engaging in the processes of scientific investigations. However, the use of activities-based curriculum does not guarantee inquiry-based instruction. This study will explore the varieties of teachers' enactment and the relations between teachers' enactment and students' learning on an activities-based curriculum. The following research questions were attempted to be answered.

Students' Learning:

(1) Were students of the target PD teacher better able to explain scientific phenomena (better conceptual understanding) compared to students of the comparison teachers?

(2) What were students' conceptual understandings of measurement and sound after doing the hands-on activities? How did they explain measurement procedures and sound phenomena?

Teachers' Enactment:

(3) How did teachers enact the curriculum? Did teachers mostly follow the written procedures of carrying out hands-on activities? Or did teachers lead students into meaningful scientific inquiries, such as describing or explaining scientific phenomena while doing hands-on activities?

Relationship between teachers' enactment and students' learning:

(4) Was there a relationship between teachers' enactment and students' learning? Did students learn more about procedures if teachers' instructions were mostly procedures? Did students learn more about scientific explanations if teachers' instructions were mostly inquiries?

1.3.2 Significance of the study

This study will bring attention to the fact that the use of activities-based curricula is not equivalent to inquiry-based instruction. Without teachers' proper guidance, students may not learn anything relevant to scientific inquiries. Teachers' enactment of the curriculum will be crucial to what students actually learned from hands-on activities. Studies regarding the recent education reform mostly focused on students' science achievement or teachers' professional development separately, but relatively few studies included how teachers actually implement those activities-based curricula in the classroom. It may seem incomplete to exclude teachers' influence while evaluating student learning of activities-based curricula. The results of this study will build on the research of teachers' enactment of activities-based curricula and the impact of teachers' enactment on student learning.

Moreover, this study will add to the research about the needs of teacher support of meaningful scientific inquiry in the curricula. The inclusion of teachers' enactment in ordinary school settings provided valuable information on the actual implementation of these curricula. Research has shown that reforms can occur under circumstances with sufficient support and resources in research settings, but more research is needed about reforms in the "real world" (Anderson & Helms, 2001). Teachers who did not participate in professional development workshops targeting scientific inquiries may lack relevant science background to utilize hands-on activities effectively. There is only limited time allocated for science classes in elementary schools. It gets more challenging for teachers to coordinate inquiry activities timely and effectively during science classes, compared to those teachers who could find support from researchers. Although there are some positive results from teachers of professional development, it is also necessary to include teachers' enactment of curricula from ordinary schools to investigate the realities of the reform outcomes.

Additionally, instead of using items selected from national exams, the assessment items in this study focused on specific learning outcomes originating from the two FOSS modules included in this study. Those nationwide standardized exams (e.g. Trends in International Mathematics and Science Study (TIMSS), published by National Center for Education Statistics) were frequently used in many large-scale studies about student achievement in science education reform. The results of these exams do provide a broad picture of student learning about science, but they fall short on describing students' own explanations of their understandings. In this study, interviews with students about performance tasks were conducted to directly explore students' conceptual understandings of targeted scientific phenomena. The inclusion of multiple assessments will provide different sources of evidence of what students learned about scientific inquiry. The use of interviews based on performance tasks not only gives students concrete contexts of the problems but also provides different ways for students to express their understanding, such as using gestures to describe how to measure an object. Furthermore, students' explanations from the performance interviews will also contribute to the literature on students' concepts of measurement and sound, which has not yet been fully established in science education especially for younger children.

1.4 ORGANIZATION OF CHAPTERS

In Chapter Two, the historical background of teaching scientific inquiry in science education is first reviewed to explain the importance of scientific inquiries from different perspectives. After the summary of current emphasis in science education, two bodies of related studies were reviewed, including the efforts in teachers' professional education and research results in student learning of activities-based curricula. In Chapter Three, the design and contexts of this study were presented as well as the planned analysis of the collected data. In Chapter Four, the results of students' learning and teachers' enactment of the curriculum were presented, followed by the exploration of the possible patterns between students' learning and teachers' enactment. In Chapter Five, the findings of the study were concluded and discussed regarding its influence and possible follow-up studies in science education.

2.0 REVIEW OF LITERATURE

The purpose of this study was to examine teachers' enactment of an activities-based science curriculum and its relation to students' scientific understandings. This review started with a brief history of the trend of engaging students with scientific inquiry in science education to explain the rationale for adopting activities-based curricula, and continued by describing valued learning outcomes in response to the current reform. After reviewing the trend of inquiry, the section that followed focused on two sources of literature: one is the reform efforts of inquiry-based instruction, and the other is the reform impacts on students' learning. The progress of the current science education reform has undertaken different aspects of education, such as teachers' professional development, curriculum development, policy making, and students' learning and assessment. Among these aspects, teachers' enactment and students' learning in the classrooms directly reflects the impact of the reform efforts.

The first source of this review involved reform efforts of curriculum design and teachers' professional development. The features of activities-based curricula and teachers' enactment of the curricula were summarized to explain the possible difficulties and needed supports while implementing the curricula. Following this vein, I introduced the theoretical background of a professional development workshop which this study was originated from.

The second source came from the research findings of student learning of scientific inquiry. Those included studies aimed at designing appropriate performance assessments to closely test students' abilities in scientific inquiry; studies making comparisons of students' learning among different types of curricula; and studies evaluating the impact of professional development by assessing students' learning.

This study examined students' understandings of two curricular modules: Measurement and Physics of Sound modules. The last section of this chapter reviewed and summarized studies about children's concepts of measurement and sound.

It should be noted that students' learning of scientific inquiry in this study was framed by the definition of scientific inquiry being chosen, the types of learning outcomes being selected, or the forms of assessments being used to represent student learning. Different researchers of hold different perspectives on each of these topics. In general, when referring to scientific inquiry in science education, it means that vocabulary and memorization of facts should be deemphasized and more in-depth thinking processes is stressed when students are engaged in inquiry-based activities. Based on this general definition of inquiry, this study was then conducted to try to answer how teachers' enactment of the curriculum could affect students' learning.

2.1 THE ROLE OF SCIENTIFIC INQUIRY IN SCIENCE EDUCATION REFORM

Science includes both process and product. The process indicates a set of procedures and logic rules that scientists employ to conduct investigations, and the product is a body of explanations and observations generated through the process (DeBoer, 2004). Both the body of scientific knowledge and the activities of scientific inquiry contribute collaboratively to understandings of the natural world and the advancement of technology. It is therefore essential to educate students in both aspects of science if we expect students to appreciate the nature of science and understand the progress and achievement of human knowledge.

The emphasis on students' abilities of doing scientific inquiries is one of the characteristics of the current science education reform (Bybee & Fuchs, 2006). "Scientific inquiry refers to the diverse ways in which scientists study the natural world and propose explanations based on the evidence derived from their work. Inquiry also refers to the activities of students in which they develop knowledge and understanding of scientific ideas, as well as an understanding of how scientists study the natural world" (NRC, 1996, p. 23). Taking a socio-cultural perspective on learning, many science educators would agree that students' learning of science should not be separated from the activities on which scientific knowledge is built (Anderson, 2002; Crawford, 2000; Dalton et al., 1997; Lave, 1988; Rogoff, 1994). Reform documents also stressed that students should engage in scientific inquiry to understand science concepts, appreciate how scientific knowledge was developed and be able to conduct scientific inquiry (NRC, 1996). The idea of teaching scientific inquiry is not new in science education, but its focus has been changed from time to time. The following section is a brief summary of the role of teaching scientific inquiry in science education.

2.1.1 Brief history of the importance of scientific inquiry in the United States

Since science was first regarded as a subject in primary and secondary school curriculum in 19th century, the importance of learning/teaching science through inquiry had been debated (Ruby, 2001). Science was considered to be different from other subjects because of the inductive-basis of scientific explanations (DeBoer, 2004). Educators, such as Herbert Spencer (1802-1903) and Johann Friedrich Herbart (1776-1841), suggested that science should be taught in its original form. They argued that scientific knowledge was constructed inductively through systematic scientific observations and discoveries; and students should be guided in their own discoveries in order to assist them in constructing meaningful understandings of science throughout the process (DeBoer, 2004). Although this notion was acknowledged by educators at the time, the notion of doing scientific inquiry in science curriculum was not widely accepted due to the comparatively large amount of time and efforts required to set up materials and equipment. Thus, textbooks were extensively used for teaching science within the limited time allocated for science class.

During the first half of the 20th century, John Dewey, an influential leader of the education movement, proposed that students' skills for critical thinking and problem solving should be emphasized in science education through inquiry teaching. Students should be prepared to ask questions, identify resources and reason about possible solutions independently. The trend in science education shifted to a more practical perspective on learning: this approach focused on preparing students for solving problems on social-related issues in addition to reasoning inductively about scientific problems (DeBoer, 2004). To help students think concretely about scientific discoveries, school laboratory activities were conducted as confirmation or illustration of textbook information (Hofstein & Lunetta, 1982) rather than focusing on reproducing scientific inquiry in class. At this time, science curriculum was still based on textbooks; laboratory work was criticized because (1) it focused narrowly on science topics and methods, (2) teachers were not teaching it effectively, and (3) its content was irrelevant to students' interests (Hofstein & Lunetta, 1982).

Beginning in the 1960s, the outgrowth of scientific discoveries and rapid economic changes urged the need to construct more robust curricula to prepare students for future scientists or science-literate citizens (Carter, 2005). The National Science Foundation (NSF) supported the development of three elementary science curricula. These curricula were consisted of hands-on investigations conducted and operated by students to create first-hand observations and regenerate scientific discoveries. The role of scientific inquiry/laboratory activities was reconsidered in the "new" science curricula, which emphasized the practice of scientific thinking skills and the understanding of the nature of science (Hofstein & Lunetta, 1982; Pine et al., 2006). Science educators were optimistic that the emphasis on scientific inquiry in these curricula would easily increase students' performance on tests (Bredderman, 1983). However, earlier development of inquiry-based curricula did not support lasting results on students' achievement (Shymansky, 1989). Thus, considering the comparatively large cost of laboratory materials and teacher efforts, most schools resumed the use of textbooks and direct instruction of scientific knowledge.

The publication of A Nation at Risk in 1983 reported crises about the quality of teaching in U.S. education and areas of poor achievement on students' test scores. This called policy makers' and educators' attention to rethinking science education reform as preparing science literacy for all students. NRC and AAAS published educational standards to inform educators about the direction and expectations of science content, teaching and professional development, student learning, and assessment. In NSES, the goals in science education were explicitly stated that students were to:

> (1) experience the richness and excitement of knowing about and understanding the natural world;
> (2) use appropriate scientific processes and principles in making personal decisions;
> (3) engage intelligently in public discourse and debate about matters of scientific and technological concern; and
> (4) increase their economic productivity through the use of the knowledge.
>
> NSES, Goals for School Science, p.13.

To better achieve these goals, it is now believed that inquiry is "central to science learning" (NSES, p.2). Inquiry in the science classroom tries to resemble the ways scientists explore the world so that students can understand the tentativeness between explanations and evidence in science, instead of confirming scientific findings written in textbooks (DeBoer, 2004). Hands-on activities designed in the reform-based curricula are intended to create environments for students to participate in scientific investigations, which mirror the way scientific knowledge is established, and to appreciate the nature of science. It is also hoped that through practicing the activities of scientific inquiries, students will be able to apply scientific thinking in everyday life to make informed decisions about science-related and societal issues.

2.2 REFORM EFFORTS IN SCIENCE EDUCATION

The notions of science education reform can be realized in different disciplines, such as curriculum development, policy making, teachers' professional development, or evaluation of student learning. At the district and school level, educational policy and curricula lead the directions and focus of education. At the classroom level, educational curricula and teachers' professional development directly influence what students actually are learning and whether students can benefit from the educational reform. In the following section, two types of studies were introduced: research on curriculum development and on teachers' professional development. To summarize the change of focus in science curricula, I compared the features and learning outcomes of activities-based curricula with traditional text-based curricula. Additionally, it might be problematic that activities-based curricula only provide little support for *full-inquiry* for teachers, especially when teachers are lacking related science background and teaching experiences (Huber & Moore, 2001). Many researchers have agreed that teachers actively interpret the intended goals of the curricula (Ben-Peretz, 1990; Clandinin & Connelly, 1992) and recognized the importance of enhancing teachers' training about inquiry-based instruction (NRC, 1996; von Secker & Lissitz, 1999). In the section which followed the introduction of activities-based curricula, I discussed teachers' role in curricular enactment and summarized the main guidelines of a professional development workshop (PPHOSS) in order to illustrate some important support for inquiry teaching.

2.2.1 Activities-based curriculum

In order to prepare students for scientific inquiry, NSF funded university and research institutes to develop several activities-based science curricula, such as Full Option Science System (FOSS), Insights, and Science and Technology for Children (STC) in elementary school science. Activities-based science curricula is also called hands-on or kit-based curricula, which featured students *doing* scientific activities such as making observations, collecting data, representing data patterns, collaborating and communicating with others. Each module of these curricula is usually packed in a big box containing common household materials, worksheets and only minimal support in the form of written texts. Students use those common materials, which can be easily related to daily life, to conduct investigations and write down their observations on worksheets. Activities-based kits require students to create first-hand phenomena and find patterns or relations from their observations, or to find the best solutions for a problem by applying appropriate scientific knowledge. For example, an activity to create first-hand phenomena would require students to mix water with salt or gravel and observe the difference between solutions and mixtures. An example of a problem-solving activity would ask students to find a way to make sound louder or travel farther. These activities were designed to exemplify the essence of what scientists do in different scientific practices so that students could learn how to do scientific inquiry in different domains of science and understand the processes of constructing scientific knowledge.

2.2.1.1 Features of activities-based curriculum

Activities-based science curricula are different from text-based curricula in several ways. First, text-based curricula focus on the descriptions of scientific facts and theories, whereas activities-based curricula are aimed at students exploring the process of scientific discoveries and generate explanations by themselves. The former emphasizes the content of science, and the later stresses the process of science.

Second, the content of text-based curricula covers a broader range of topics, while activities-based curricula spend more time (6-8 weeks) on each topic so that students could explore the discovery process of scientific phenomena and theories (Sherin, Edelson & Brown, 2004). Thus, fewer topics are contained in activities-based curricula. But the curricula emphasize greater depth of understanding by engaging students in the discovery processes of different phenomena around the same topic.

Third, text-based curricula usually present science as abstract statements of theories, whereas activities-based curricula create meaningful contexts so that the learning of science concepts and processes occur naturally through solving problems or conducting investigations. The problem with teaching abstract theories is that students may have difficulty applying science knowledge in other contexts or maintaining their interest in science (Warwick & Stephenson, 2002). Those facts and methods are not useful to students until they know in what situation to apply the knowledge.

Generally speaking, activities-based curricula are based on the idea that scientific knowledge should be learned from doing scientific inquiry; thus, only minimal support for students is provided from written texts. Students gain their understandings of the content of science through the process of hands-on activities. While engaging in the activities of scientific inquiry, students are expected to learn a lot of things other than just carrying out the routines of conducting investigations. For example, students have to construct scientific concepts from various activities throughout and across the modules, communicate ideas with others, use evidence to support their ideas, build understandings about how scientific knowledge is established, work collaboratively with peers, etc (Marx et al., 2004). It is thus crucial that teachers be aware of the rich opportunities embedded within hands-on activities and guide students to understandings and appreciations of scientific inquiry.

2.2.1.2 Possible problems with activities-based science curricula

The design of the curriculum can be important to teachers' implementation in the classroom. Taking FOSS curriculum as an example, the teacher's manual is consisted of five sections, including *overview*, *materials*, *investigations*, *assessment*, and *other resources*. Among these five sections, the investigations section always contains the most pages regarding the preparations and procedures of setting up hands-on activities. These instructions are listed as a cooking recipe detailed with step-by-step procedures illustrating how to craft the materials and proceed with the activities. Along with these procedures are suggestions for class management and suggested discussion questions. Those related science concepts regarding the meanings of the activities are included implicitly in a separate overview section. It is possible that while teachers are careful about following each step of the procedures, they may not intentionally incorporate the underlying concepts with the ongoing activities in their instruction. Hands-on activities do not reflect any aspect of scientific inquiry if students are simply executing laboratory routines such as crafting materials, manipulating equipments and recording results.

In math education, the emphasis on computational procedures or conceptual understanding in instruction has also been a topic open for debate. The practices of procedures can be characterized as the skills or actions required while performing arithmetic calculations (Rittle-Johnson, Siegler, & Alibali, 2001); whereas conceptual understanding involves the justifications and explanations of the execution of such activities (Niemi, 1996). Taking fraction as an example, students who were able to perform computational procedures of fraction might not understand concepts of fraction properly. For instance, students might view fractions as symbolic entities, rather than concepts or principles, have difficulties to explain fraction problems conceptually, or have trouble to evaluate or justify their problem-solving procedures (Niemi, 1996). While procedural knowledge may be essential to learning fractions, conceptual understandings are also important to making sense of the ongoing procedures or activities. Similar learning problem might occur in activities-based curriculum. Procedural practices of computation or scientific investigations do not come with conceptual understanding. The separation of experimental procedures from related conceptual background in the curriculum may lead to the lack of science concepts in teachers' instruction and leave out opportunities of scientific thinking practices. It needs to be cautious in science education that the imbalance of procedural and conceptual understanding in instruction is likely to result in students not learning science ideas at all. The conceptual knowledge hidden behind the execution of hands-on activities requires teachers to press students for higher level of thinking and understanding toward scientific inquiry. The organization of teacher's manual does not provide much support for teachers to help students develop ideas throughout the investigation processes. Teachers' professional background of pedagogy and scientific inquiry are influential to how hands-on science is implemented in class, which will be addressed in the next section.

2.2.2 Teachers' role in educational reform

Moving from text-based instruction to inquiry-based instruction, teachers are facing challenges of updating themselves with new ideas in educational reform and integrating new curricula into their practices. There are significant changes from the traditional way of teacher-centered instruction, where teachers lecture students in large-group contexts, to a more student-centered instruction, where students work on scientific projects in small groups to conduct investigations or solve problems. Lecture-based instruction displays the product of scientific discoveries, whereas inquiry-based instruction puts emphasis on the process of scientific discoveries. In reform classrooms, while students are involved in scientific investigations, teachers are responsible for utilizing opportunities for scientific inquiries and guiding students through the construction of scientific knowledge. Consequently, teachers are crucial to the achievement of the education reform (von Secker & Lissitz, 1999).

The challenges of science educational reform also bring up the issue of equity (Lynch, 2000). Teachers' readiness for reform-based curricula may directly reflect on their teaching, and teachers who are less experienced and hence less prepared for educational reform are usually the majority who teach at schools with diverse students background (e.g., SES or ethnicity). Students with lower SES background or from minority ethnic groups are more likely to be in classes of teachers who have less experience in teaching and have less content knowledge about science. It may also lead to problems regarding educational resources available to teachers and students to properly adapt to educational reform.

In order to apply reform-based instruction properly to enhance students' deeper thinking, teachers have to acquire new knowledge and teaching strategies (Borko & Putnam, 1996; Schneider, Krajcik & Blumenfeld, 2005). Support from professional development is essential and necessary for the success of education reform (Schneider et al., 2005). In the next section, descriptions of a teachers' professional development workshop was summarized to explain the kind of teacher support is needed for activities-based curricula.

2.2.2.1 Support from professional development: Pittsburgh Public Hands-On Science Study (PPHOSS)

The planned curricula designed by curriculum developers still require teachers actively interpret the intended learning goals and interact with students to carry out the anticipated learning processes. Teachers do not simply act as an implementer of the curriculum (Ben-Peretz, 1990; Clandinin & Connelly, 1992), but constantly modify and redesign the curriculum to make tasks or activities appropriate for student needs or school contexts (Remillard, 2005). Since it is inevitable for teachers to make necessary changes, support from professional development is important to ensure teachers' proper enactment of the curriculum.

Pittsburgh Public Hands-On Science Study (PPHOSS) is a teacher professional development workshop, which aimed at promoting effective science teaching. PPHOSS took the form that researchers collaborated with in-service teachers as a community to support teachers' professional development (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003). Cartier (2003) worked with six urban public school teachers collaboratively on supporting teachers' enactment of FOSS curriculum by introducing Big Ideas, Tools, and Talk in teachers' plan of instructions. During this professional development workshop, teachers first identified <u>Big Ideas</u> (or key conceptual models) of a teaching unit. Big Ideas are simple but powerful science concepts which students can use to explain things within and across various activities so that students can build and develop other related science concepts later on. For example, in an Insects module, one big idea is *living things go through different stages of a life cycle*. As students observed different insects, the same big idea reoccurred and could be revisited to discuss the similarities/differences for different insects. This big idea can later be connected with other living things at different grade levels.

Teachers are encouraged to use <u>Tools</u> (such as tables, graphs, illustrations or object models) to summarize, organize or represent data patterns for students to use and explore possible relations among the observed data. The use of tools was aimed to help students represent data or ideas in multiple ways to facilitate the discovery of patterns or reconstruct their ideas. Tools can serve different purposes such as being used as manipulative tools, experiment tools, representation tools. The kind of tools encouraged in PPHOSS was mainly referred to representations of data patterns, concepts or ideas with the purpose of supporting student explanations.

Lastly, teachers engage students in Talk to describe observations, communicate possible data patterns, explain ideas to others, or support their arguments using related evidence. Teachers can also monitor students' understandings of the lesson or science through engaging students in explaining their ideas or discussing with others. These three elements may represent the essence of inquiry-based instruction and thus prepare students for authentic scientific inquiries.

The theoretical framework of using Big Ideas, Tools and Talk (BITT) as critical elements of inquiry-based instruction is based on two sources of research. Researchers reviewed the history and philosophy of science and suggested that perhaps the most important goal of scientific inquiries is to explain the mysterious world (Salmon, 1998). Searching for explanations can serve a purposeful meaning for scientific activities in elementary science. While applying authentic scientific practices in the science classroom provides a meaningful context for students, it may be inappropriate and unnecessary to overwhelm students with complicated experimentation techniques as scientists do in their practices. The essential goal of explanations in science should play a central role in making sense of scientific investigations. The establishment of reasonable explanations requires students to construct conceptual knowledge based on evaluating the data patterns they observed and its consistency with several plausible explanations. Explanations seem to be useful to engage students in reasoning scientific ideas, constructing their understandings, and communicating with others. Thus, an attempt to explain what was observed using basic and simple ideas can be one of the keys to helping students understand both scientific inquiries and science concepts.

Another source of research for BITT came from studies in educational settings. Cartier, Passmore and Stewart (2001) introduced a practice framework in the absence of a more general framework of scientific practices in science classrooms. Their framework included the essential elements of inquiry across different scientific disciplines. This practice framework illustrated that scientific inquiries involve processes of observing data, finding possible patterns of data, looking for explanations to account for current data, and evaluating for best explanations considering all data. By adopting this framework as guidelines in an evolutionary biology classroom, students learned to make observations, explain data patterns, evaluate and modify their working hypotheses (Cartier, Barton & Mesmer, 2001). The basic structure of BITT was incorporated from this practice framework shown in Figure 1. According to this model of effective science teaching, teachers focus on *Big Ideas* while directing students to their observations, looking for data patterns, and searching for explanations. *Tools* and *Talk* are used as pedagogical supplements in the classroom. The use of models or representations (*Tools*) in science classrooms, such as data charts, tables, models, manipulative materials, could enhance students' understandings and communication of science ideas (Cartier, 2003; Lehrer & Schauble, 2002). Through students' talk about constructing explanations and supporting claims, teachers are able to observe student ideas and guide student thinking. PPHOSS introduced BITT into the planning of instruction to support teachers to structure hands-on activities into inquiry-based instruction, so that students would be systematically guided toward understandings of scientific inquiries and scientific explanations.

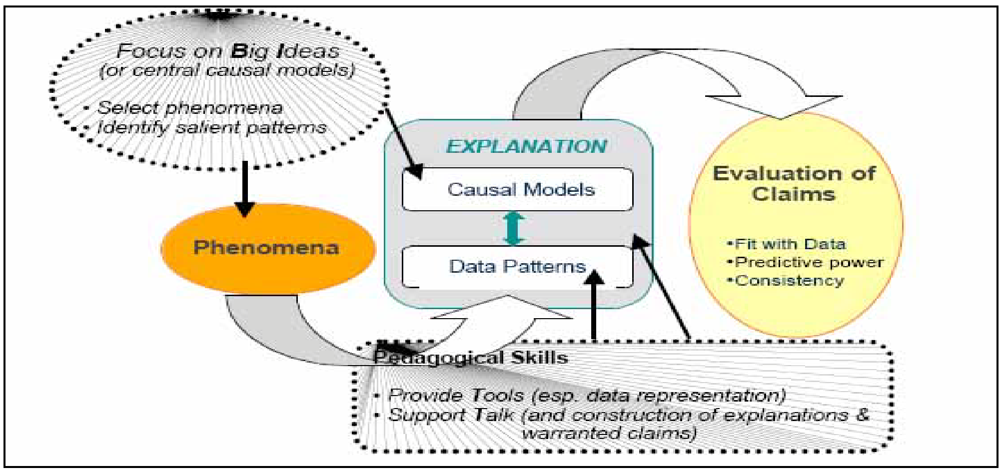


Figure 1. BITT model of effective science teaching (Cartier, 2003).

2.3 REFORM OUTCOMES OF STUDENT LEARNING

Research suggests that students gain deeper understandings by actively engaging in activities consisting of critical thinking and problem solving within meaningful contexts (von Secker & Lissitz, 1999). However, how teachers should engage students in higher-order thinking activities within scientific inquiries and how students can benefit from scientific inquiries are still open for discussion. Among recent reform efforts, the evaluation of students' learning outcomes can be categorized into three aspects: studies which focused on the development of appropriate educational assessments targeting inquiry skills; studies which included specific learning outcomes as evidence of the impact of particular professional development workshops; and studies which examined the advantages of adopting activities-based curricula or compared the differences with text-based curricula. The following section provides a few examples of each type of study. This summary was not meant to be exclusive to the detailed results of each study, but to illustrate different attempts of examining students' knowledge of scientific inquiry and science contents.

2.3.1 The development of appropriate assessments

Since achievement in science is no longer viewed as simply the recall of factual knowledge, traditional assessments, such as multiple-choice items, may not be representative of or comprehensive enough for testing students' abilities regarding conceptual understanding, problem solving and scientific inquiries (Baxter, Shavelson, Goldman, & Pine, 1992; Shavelson, Baxter, & Pine, 1991). Alternative assessments provide a closer look at students' abilities for doing scientific inquiries while encountering a specific science-related problem. Alternative assessments such as performance assessments, which require students to manipulate tools and conduct investigations to solve a problem, provide a meaningful context for students to demonstrate their reasoning based on actual experimental or observational results rather than descriptions from an assessment item.

Researchers such as Shavelson, Baxter, Goldman, Pine, and Solano-Flores developed several performance assessments in science. One of the performance assessment tasks developed by Shavelson and Baxter (1992) required students to conduct an experiment to determine which one of the three paper towels soaked up the most and the least water. Such performance tasks usually provide a context for students to solve a predetermined problem. Students are asked to plan and conduct an investigation, demonstrate the use of equipment, and explain how to solve the problem. Since these assessment items are intended to measure inquiry abilities, content-specific knowledge is not heavily related to the contexts of the problem.

However, there remain challenges for the development of performance assessments (e.g., Solano-Flores & Shavelson, 1997). Although the abilities required in completing the performance tasks seemed to be more consistent with those provided in the science education standards, students' performance often varies from one task to another or from one content area to another. It is difficult to find representative tasks to cover scientific practices across different disciplines. Therefore, multiple tasks were recommended by researchers in order to generalize from students' observed performances of the tasks to students' abilities of doing scientific inquiry (Shavelson et al., 1991).

Studies which were designed to compare the results between performance assessments and multiple-choice tests implied that these two types of assessments may be testing different abilities (Shavelson et al., 1991). Shavelson, Baxter and Pine (1992) compared over 300 fifth- and sixth-grade students in two school districts on students' outcomes using performance assessments and multiple-choice tests. Three performance assessments (i.e., paper towels, electric mysteries and bugs) were compared with a standardized, multiple-choice science achievement test (i.e., Comprehensive Test of Basic Skills). They found that the correlations between the two tests were only moderate, and thus suggesting that different aspects of abilities were measured in these two types of assessments.

The development of appropriate educational assessments focusing on skills of scientific inquiry could lead teachers to teach what is being assessed and thus emphasize skills required for conducting scientific investigations rather than teaching strategies for locating the correct multiple-choice answer (Shavelson et al., 1991). Adopting appropriate educational assessments which are consistent with the science standards is important to ensure the implementation of science education reform and evaluate the outcomes of the reform.

2.3.2 Comparison between text-based curricula and activities-based curricula

Another kind of research aimed at documenting the impact of educational reform on students' learning involves making comparisons between activities-based curricula and text-based curricula (e.g., McCarthy, 2005; Pine et al., 2006). Students' learning outcomes after instruction using different curricula seem to provide a direct comparison regarding the effect of the two kinds of curricula. However, studies aimed at such comparison were mostly drawn from larger studies which provided professional support for teachers using activities-based curricula (e.g., Dalton et al., 1997; Schneider et al., 2002). These studies' findings consistently that students learned better from activities-based curricula (McCarthy, 2005; Schneider et al., 2002). Nevertheless, it is not clear whether students from regular classes whose teachers were not provided with or did not seek professional support regarding scientific inquiry could gain significantly from activities-based curricula.

Among the few studies which assessed student learning from regular classes with no professional support provided, the results did not show consistent evidence supporting the use of inquiry-based science curriculum. For example, Pine et al. (2006) examined students' learning outcomes of using either hands-on or text-based science curricula from nine schools for about 1000 students. In order to consider the factor that procedural knowledge (i.e., scientific inquiry) should not be tested by paper-and-pencil tests, they used four performance assessments representing different science fields and compared students who were taught with different types of curricula. After controlling factors such as socioeconomic status and students' cognitive ability, the results indicated that students' performance in the tasks was not different between the types of curricula being used.

The failure to find positive correlations between activities-based curricula and better student performance in regular classes may be due to the difficulty of designing fair tests to compare the two types of curricula. Activities-based and text-based curricula emphasized different skills and content knowledge. Students who were taught with text-based curricula were exposed to short science facts across a broader range of topics; whereas students taught with activities-based curricula spent more time with fewer topics and focused on the processes of conducting scientific investigations (Pine et al., 2006). Additionally, the lack of significant results might be partly caused by teachers from regular classes not being well-prepared for activities-based curricula (Huber & Moore, 2001) or the reform curricula not having immediate impact on students' learning until later (Marx et al., 2004). In summary, one might conclude that under the circumstance of adequate teacher support, students can benefit from the use of activities-based curriculum and perform better in their skills of scientific inquiry. It still requires further exploration in future studies to examine closely the impact of different types of curricula, the impact of inquiry teaching, and the impact of teacher enactment of the curricula on students' science achievement.

Another line of studies which also suggested positive relationship between hands-on science curricula and better student performance focused on non-mainstream students such as students with learning disabilities (Cawley & Cawley, 1994; Dalton et al., 1997; McCarthy, 2005; Scruggs & Mastropieri, 1994), students who speak English as foreign language (Lee, Buxton, Lewis, & LeRoy, 2006), or students from diverse cultural backgrounds (Lee et al., 2006). These studies usually involved teacher intervention or professional development program and the results provided strong evidence that students with special needs were better able to learn science by doing it. Students who were taught with activities-based curricula performed better in their understanding of science content using short-answer assessments, in the number of ways they generated to conduct science investigations (McCarthy, 2005), and in guided elicitations for designing an investigation (Lee et al., 2006). The use of activities-based curricula seemed to provide students with learning disabilities students with concrete experiences which strengthen the meaningfulness of the scientific knowledge to be learned (Mastropieri & Scruggs, 1994).

2.3.3 Evaluation of professional development

Students' learning outcomes were also tested in studies which concerned about the impact of teachers' implementation of the curriculum on students' learning when teachers were provided with supports through professional development or educative curriculum (e.g. Dalton et al, 1997; Raghavan, Cohen-Regev, & Strobel, 2001; Hapgood, Magnusson, & Palincsar, 2004). These professional development efforts were aimed at supporting inquiry-based learning by working with in-service teachers as a research community, by enhancing teachers' content and pedagogical knowledge, or by revising curricula to better support teachers' instruction. In these studies, students' learning outcomes were usually based on selected items from standardized tests (e.g. Trend in International Mathematics and Science Study [TIMSS]). These nationwide assessments consisted mostly of multiple-choice items which went beyond superficial recall of facts. However, comparing students' learning outcomes on selected test items may leave out differences directly related to the specific concepts taught in the module (Marx et al., 2004; Porter & Smithson, 2001). National exams usually cover a wide variety of topics with only a few items representing each topic. The selected items from national exams might not thoroughly assess students' understandings around a specific topic. Some relevant concepts of the specific topic might be left out from such large scale studies. The inclusion of various important concepts which directly relate to the target modules is necessary for complete exploration of the differences in students' learning outcomes.

To better assess students' learning outcomes, this study designed and used assessment items which were to examine thoroughly the focused science content (measurement and sound). The next section reviewed studies of students' concepts about the target science modules taught by the teachers in this study to further explain the understanding of elementary science concepts covered in this study.

2.4 SCIENCE CONTENT COVERED IN THIS STUDY

Participants in this study were elementary third grade students and teachers. This study collected students' learning outcomes on two third-grade modules of the FOSS curriculum: Measurement and Sound modules. Children's conceptual understandings of these two topics was summarized and explained in the following section.

2.4.1 Children's development of measurement concepts and skills

Measurement is fundamental in both mathematics and science and can be easily applied in everyday life. The development of measurement concepts may start by counting distinct items at about age two, followed by the ability to quantify continuous attributes (such as length) (Barrett & Clements, 2003). Piaget, Inhelder, & Szeminska (1960) suggested that children's development of length measurement comes from visual comparison, then putting things side by side for comparison. The concept of using a measuring tool to provide a common measure occurs even later. Although children may have experiences of watching others using measurement tools and thus learn how to execute measurement procedures, the development of measurement concepts take longer to be fully appreciated (Barrett & Clements, 2003; Clements, 1999a, 1999b; Hiebert, 1984; Joram, Subrahmanyam, & Gelman, 1998; Lehrer, 2003; Stephen & Clements, 2003).

There are at least five concepts children have to comprehend and coordinate throughout the process of development (see also Lehrer, 2003; Stephen & Clements, 2003 for a complete inclusion of measurement concepts).

1. One has to recognize the relation between the property to be measured and measurement units which represent the property of interest. Some properties such as length may seem more obvious than others such as area or angle, which children often misrepresent using length (Lehrer, 2003). The correspondence between a property and its related measurement tool(s) has to be established appropriately for students to fully understand the reasoning of which tool to use.

2. Measurement units have to fully cover the property to be measured without leaving any space. For example, while measuring the length of a book with blocks, one end of a block has to be placed closely to one end of another block so that the entire length is covered. Younger children might not be aware of the inaccurate results caused by leaving cracks between the blocks and do not pay sufficient attention to the placement of units closely to each other while measuring (Lehrer, 2003).

3. Measurement units can be used iteratively and will not run out. If the measured length is longer than the available measurement units (e.g. the length of a ruler), one can mark the end of a ruler, move the ruler to align with the mark and proceed using the ruler. Younger children often align one end of the measurement tool and say that the tool is too small to measure (Kamii & Clark, 1997).

4. Within the same measure, measurement units have to be identical. It is very common among younger children to combine the use of different measurement units within the same measure. Children are not aware of the problems such as not being able to reproduce results or communicate with others due to the use of different units. In this paper, this particular concept is also called "concept of uniformity" which indicates the understanding of uniformed measurement units.

5. The property to be measured (e.g. length) remains the same regardless of different ways of measurement, such as using measurement units of different sizes, or not being placed to align with the zero-point of a ruler. Children should come to an understanding that since the property being measured remains unchanged, using longer measurement units should result in a smaller number of measurement results,

and not aligning with zero-point on a ruler does not make the measured property longer.

Additionally, the hands-on activities suggested in the FOSS curriculum (see Appendix A) emphasize the practice of measuring procedures, but these measurement activities are not aimed at any purposeful investigations. For example, students were asked to measure their body parts using measuring tape. But such measuring activities did not lead to the discovery of scientific phenomena or to answering research questions. Without a meaningful purpose for doing hands-on activities, teachers may just follow the stepwise instruction to make sure that the correct use of measurement tools would be the only purpose of the entire module.

Although the purposes and concepts of measurement are essential to students' correct use of measurement tools, they are not necessarily required for the proficient use of tools. It is rare to see measurement concepts being emphasized in a curriculum. In this study, we differentiated students' learning about measurement procedures and concepts in order to distinguish those who are skilled in measuring and those who comprehend the basic concepts and understand the purposes of measurement.

2.4.2 Children's development of the concepts of sound

Sound is one form of energy that relates closely to our everyday experiences. Some interesting phenomena of sound can be easily observed and connected, such as echoing, sound isolation, or sounds from unique musical instruments. Sound is caused by vibrations of objects which have been hit by a force. The vibrations are transmitted from one place to another through the compression and dilation of the media. The energy travels to the furthest end of the vibrations, but within the medium, the matter of the medium only oscillates back and forth instead of moving with the energy to other places. It is important to note that sound is a process of energy transmission and does not share the physical properties of objects, such as substantiality (objects take up space but not energy).

Students often come to school with numerous experiences of sound. According to Mazens and Lautrey's (2003) review, Piaget (1971) reported that younger children think that nothing passes through ears when one hears something, or that sound stays inside of an object and comes out of the object when we hear sound. Seven- to eleven-year-old children start to conceive sound as air or describe that sound as being spread out in all directions through air. Similar to other concepts of energy (e.g. heat), students often have difficulties understanding sound as processes of energy transmission instead of properties of objects (Lautrey & Mazens, 2004). Such misconception exists among children (Driver, Squires, Rushworth, & Wood-Robinson, 1994; Lautrey & Mazens, 2004; Mazens & Lautrey, 2003) and some college students (Linder, 1993; Linder & Erickson, 1989). As children grow older, they gradually discard some properties of matter in their concepts of sound. Mazens and Lautrey (2003) analyzed 5- to 11-year-old children's development of concepts of sound based on their justifications for four situations regarding different properties of matter, including substantiality, trajectory, weight and permanence. The substantiality problem required students to explain why we could still hear sound from the other side of a wall. The trajectory problem required students to draw the paths sound travels when a few people were around a sound source. The weight problem was regarding the misconception that objects would lose weight slightly when they made sound. The permanence problem required students to explain how far they think sound could travel. The results showed that older children abandoned properties of matter as attributes of sound in the order of weight, permanence, and substantiality. The property of substantiality was the most difficult one to discard. For example, children rationalized that there were holes or cracks in the obstacles between the trajectory of the sound source and ears so that sound could sneak out and be heard. In contrast, an explanation without considering sound as a substance is that sound (like other forms of energy) travels by passing vibrations through mediums.

Even for some college students, the conception of understanding sound as matter may still occur in their explanations of sound. Their explanations of the factors affecting the speed of sound still revealed their perspectives of sound as an entity carried by molecules or transferred between molecules (Linder & Erickson, 1989). Some college students also explained that sound traveled slower in wood than in the air because wood provided more resistance (Linder, 1993). They might adopt physics terminologies to explain sound, but that does not guarantee they fully grasp the idea of sound as energy (Linder, 1993). Students' concepts of sound have not yet been studied extensively in educational research compared to other topics in physics such as force and motion, electricity or energy. Some concepts of sound might still be challenging for older children or even elementary school teachers to understand.

In the FOSS Physics of Sound module, students are involved with activities such as exploring and distinguishing different sounds made by different objects, observing objects vibrating when they were making sound, identifying the relations between pitch and the length of an object, and listening to sound as it travels through water, air and wood materials (see Appendix B). These activities seem to focus on observing the properties and the processes of sound. However, similar to the Measurement module, these observations are not followed by explanations of the phenomena in order to merge into meaningful scientific inquiry. Students were doing isolated hands-on activities, but these activities did not blend in any scientific thinking or relate to students' daily life if teachers fail to make explicit connections for students. Moreover, the lack of such important connection may not be seen by assessing fact-based knowledge such as the relation between the pitch of sound and the length of objects. Thus, in this study, the distinction between students' explanations of the processes of sound and students' observations of the pitch of sound was made while examining students' understanding of sound.

3.0 METHODOLOGY

3.1 DESIGN OF THE STUDY

The purpose of this investigation was to (1) study teachers' enactment of hands-on science curriculum with or without supports from professional development of PPHOSS; (2) evaluate the impact of PPHOSS on students' learning; and (3) explore the relations between teachers' enactment and students' learning. The design of this study was based on a quasi-experimental design with non-equivalent groups and posttests only. The classroom taught by the target teacher who participated in PPHOSS was the target group, and the classrooms of two other teachers who did not participate in PPHOSS were in the comparison groups. These three classrooms were from different schools whose teachers had somewhat similar teaching backgrounds (as shown in Table 1). Students' learning was collected after the instruction of Measurement module and Physics of Sound module.

3.2 PARTICIPANTS

Three teachers in this study were from a large urban school district in a Midwestern city of the United States. The target teacher, Sarah (a pseudonym), had been involved in PPHOSS for two years and was selected in this study due to her availability among other PPHOSS teachers[1].

[1] There were six participating teachers in PPHOSS workshop. At the time this study was conducted, two teachers were not assigned to teach science in that particular school year; one teacher was assigned to teach K-5 grades, which was a major change from her previous teaching duty (fourth grade science only); two teachers from lower grade level (K and first grade) were not selected considering younger children being less experienced in expressing their ideas clearly using written assessments and interviews.

Sarah had about five years of teaching experience in science in an urban public school and taught third-grade science and math in a self-contained classroom.

Two comparison teachers were identified by their district supervisors based on the following criteria: (1) their implementation of FOSS curriculum was considered as consistent and faithful by their supervisors, (2) their teaching expertise in science was similar to the target teacher, and (3) the demographics of their schools were similar to the school of the target teacher. These two teachers, Clark and Margaret, were both science specialist teachers who taught science across different grade levels of in their schools. The demographic information for the three teachers and their schools was summarized in Table 1. Although we tried to find comparison teachers whose schools were consisted of similar demographic background as the target teacher, it was always difficult for educational studies to recruit participating teachers with expected background. In this study, only one of the comparison teachers had similar school demographics as the target teacher. The school demographics of the other teacher, Margaret, were slightly different from the target teacher. The possible impact of schools' different demographics on student learning will be addressed in the discussion section of this document. Additionally, the teaching responsibilities of the three participating teachers were not entirely the same while this study was conducted. The target teacher, Sarah, taught math and science in a self-contained classroom, whereas the comparison teachers, Clark and Margaret, were both science specialist teachers who taught from kindergarten through fifth grade of the entire school.

Table 1. Demographics of the participating teachers and schools

	Target teacher (Sarah)	Comparison teacher 1 (Clark)	Comparison teacher 2 (Margaret)
Teaching experience in science	5 years	7 years	8 years
School demographics	97% African American 1% White	94% African American 2% White	37% African American 57% White
Parent socioeconomic status (school-based)	89% free lunch 5% reduced lunch	93% free lunch 1% reduced	55% free lunch 7% reduced
Participating students	13 out of 50	12 out of 50	16 out of 25

3.3 CONTEXTS FOR PROFESSIONAL DEVELOPMENT

During the professional development workshop, PPHOSS, researchers worked collaboratively with in-service teachers as a research team to construct and facilitate the implementation of effective science teaching. Figure 2 shows the timeline of the workshop and when the data collection of this study occurred. During the workshop, six participating teachers met for a week with researchers during summers and every month during the academic years. Teachers were first involved in various group activities to mirror or experience scientific inquiry. For example, teachers were engaged in a "cartoon activity" (Cartier, Passmore, Stewart, & Willauer, 2005) in which teachers were provided with several cartoon snapshots and asked to tell a story by rearranging or removing cartoon clips. After the activity, participating teachers reflected on its similarities to scientific practices, such as finding a coherent story (theory) for existing clips (data), or the construction of a story can be biased by prior knowledge. Teachers were then introduced to Big Ideas, Tools and Talk (BITT) model and discussed the use of the BITT model as support for hands-on activities. During the second year of the workshop, each participating teacher chose a target FOSS module appropriate for their teaching grade and worked collaboratively with the researchers on the implementation of BITT. They were involved with practices to identify central scientific concepts underlying related hands-on activities within the same module, to utilize or create useful tools and to engage students in conversations to assist students constructing/reasoning key concepts regarding first-hand observations produced with FOSS activities. At the end of the workshop, each teacher presented their lesson plans of their target modules to the entire research team with the implementation of the BITT model and discussions regarding possible connections across different modules.

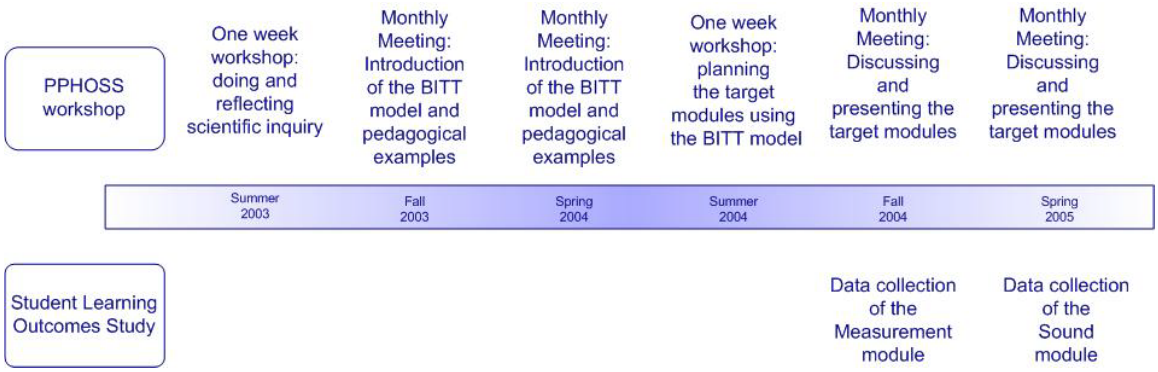


Figure 2. Timeline of the PPHOSS workshop and the time of data collection of this study.

3.4 DATA SOURCES

Data were collected for the Measurement Module (during the fall of 2004) and Physics of Sound Module (during the spring of 2005) in the classes of the target teacher and two comparison teachers. Three types of data were collected for each Module, including classroom observations during the instruction period, written assessments and student performance interviews for each participating student after each Module was completed.

3.4.1 Classroom observations

Three class observations were scheduled for each teacher within each module[2] to document how teachers implemented FOSS in their own practices, particularly how they introduced science concepts embedded in the hands-on activities. The selection of classes being observed was based on teachers' lesson plans and convenience and may not focus on the same curricular activities among the three teachers. The observations focused on how teachers carried out hands-on activities and whether teachers made conceptual connections between hands-on activities and the underlying big ideas. For example, teachers' instructions that involved the rationale for conducting hands-on activities, or explanations of data patterns or investigated results would be

[2] Class observations of Clark were only collected for Measurement module but not for Sound module because Clark was ill throughout the teaching of Sound module and the substitute teacher was hesitant to be observed for this study.

considered as making conceptual connections. A general picture of teachers' enactment of the curriculum and classroom culture was to be illustrated in the observations.

3.4.2 Student written assessments

The written assessments used in this study included multiple-choice and short-answer items. The items for written assessments were selected or designed based on the concept maps regarding the identified Big Ideas and expected learning outcomes of each module (Figure 3 & 4). The items were selected from FOSS *end-of-module assessment,* which the school district currently used as their evaluation tools in science achievement. The FOSS assessment consisted of written assessment and performance assessment. Only some of the FOSS assessment items were chosen. These items coincided with the learning outcomes we identified in the modules and did not depend on students' memorization of science facts.

In order to better assess student learning of the measurement concepts and skills identified in this study, we developed another written assessment called Big Ideas (BI) written assessment (see Appendix C and D for the test items). This assessment was designed to highlight the fundamental concepts which were implicit in the FOSS module, but missing from the FOSS assessment. The BI assessment included multiple-choice items followed by short-answer items for student justifications. For example, one item was to identify a tool to measure a jump rope using same-sized pencils, different-length crayons, and different-sized wood blocks. It was followed by another question asking which tool they thought was better and why. Short-answer items in the written assessment were scored by two graduate students. The disagreements were resolved through discussion. The inter-rater reliability for the written assessment of the measurement and sound module is 94% and 95% respectively.

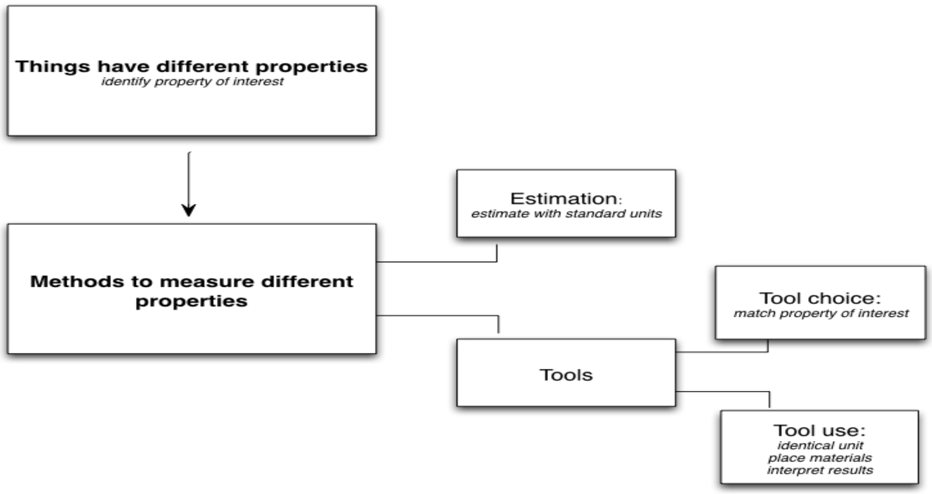


Figure 3. Identified Big Ideas and the expected learning outcomes of the measurement module.

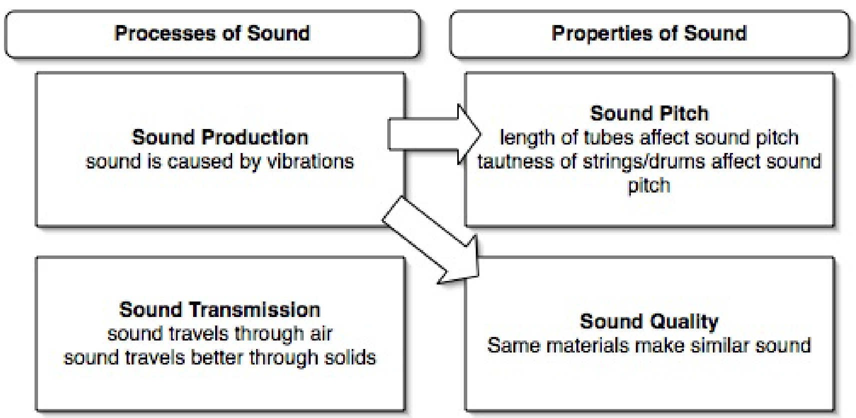


Figure 4. Identified Big Ideas and the expected learning outcomes of the physics of sound module

3.4.3 Student interviews

Written assessments can be challenging for younger children due to their lack of proficiency in reading and writing. In this study, one-on-one student interviews were conducted by two graduate students to explore students' explanations of scientific concepts in greater depth. Each interview took about 20 minutes in the form of performance assessment in which students were presented with situations of everyday problems.

3.4.3.1 Performance tasks for the measurement module

Two performance tasks were designed for the student interview. Table 2 illustrated the expected learning outcomes identified in this study and its corresponding assessment items in either selected items of FOSS assessment, BI assessment or performance interview. The first task of the interview was started by asking students to estimate how long a book is and to measure with a ruler. Students were then told that the interviewer wanted to compare the size of two books, one shown to students and one owned by “Susan”. Students had to first decide what attributes to measure about a book. Given that Susan did not have a ruler and lived at a distance, students then had to select among three tools that both the interviewer and Susan had at hand. Students could choose from AA batteries, broken crayons and different-sized screws and demonstrated how they would use the tool to measure. This task targeted students’ measurement concepts of making reasonable estimation and awareness of the need to use identical units of measurement, and students’ measurement skills of length.

The second task required students to separate dog food and water so that two dogs would have the same amount and not get jealous of each other. Students selected tools and demonstrated their use of the tools. This task was intended to see if students identified appropriate attributes to measure about dog food and water and if students can use those tools appropriately to measure weight and volume.

At the end of the student interview, students were asked to explain further about an item taken from the BI written assessment related to the concept of uniformity. This written assessment item presented three ways to measure a jump rope by using same-length pencils, different-length crayons and different-length blocks. Students had to explain which way was better to the interviewer.

Table 2. Test specifications of the measurement module.

<table>
<tr><th colspan="2"></th><th colspan="2">Expected Learning Outcome</th><th>Selected FOSS written assessment</th><th>Big Idea written assessment</th><th>Performance Interview</th></tr>
<tr><td colspan="2">Things have different properties</td><td colspan="2">Ability to identify properties of interest</td><td></td><td>Identify and explain width is the property of interest (2a)</td><td>Identify length/width/ thickness is the property of interest (1d; 1e)
Identify weight/volume is the property of interest (2)</td></tr>
<tr><td rowspan="5">Methods to measure different properties</td><td>Estimation</td><td colspan="2">Estimate using standard units</td><td>Estimate length of a piece of paper (3)
Estimate mass of a piece of paper (11)
Estimate the height of an adult (15)</td><td></td><td>Estimate the length of a book (1b)</td></tr>
<tr><td rowspan="4">Tools</td><td>Tool Choice</td><td>Select a tool that matches property of interest</td><td>Identify an appropriate tool to measure temperature (4)</td><td>Select a tool to measure width (2c)
Select a tool to measure length (3b)</td><td>Select a tool to measure length (1c)
Select a standard tool to measure length (1d)
Select a tool to measure weight/volume (2)</td></tr>
<tr><td rowspan="3">Tool Use</td><td>Measure with identical units</td><td></td><td>Identify measurement tools with identical units (3a, 3b)</td><td>Identify measurement tools with identical units (1e)</td></tr>
<tr><td>Place materials appropriately</td><td></td><td>Draw a balanced scale with equal numbers of marbles on both sides (1b)</td><td>Measure the length of a book (1c)
Separate dog food and water into halves (2)</td></tr>
<tr><td>Report results appropriately</td><td>Predict an object’s weight on a balance scale (14)
Read the result from a graduated cylinder (26)
Compare the lengths of two nails (28a, 28b)
Read the result from a thermometer (24)</td><td>Predict the result of a scale with unequal weights (1a; 1b)</td><td>Measure the length of a book with non-standard tool (1e)
Separate dog food and water into halves (2)</td></tr>
</table>

3.4.3.2 Performance tasks for physics of the sound module

Three tasks were conducted in the student interview. Table 3 summarized the identified learning outcomes with the related assessment items in written assessments and performance tasks.

<u>Task 1</u>. During the interview, students were first presented with a sound (a phone ring) but could not see the sound source. They were prompted to explain how sound traveled from the source to their ears with questions such as: “How do you know what happened if you couldn’t see it?”; “How does sound get to your ears?” Then students were presented with two sets of cups. One set of cups consisted of two plastic cups which were connected with a string at the bottom of both cups, and the other set of cups also included two same plastic cups but without a string in between. Students had to predict and explain which set of cups could be heard better if a cell phone is covered by one cup and someone listens from the other cup. This task targeted students’ concepts of sound transmission by providing them with two phenomena of sound travels through air or string.

Table 3. Test specifications of the physics of sound module

		Expected learning outcomes: student explanations	**Selected FOSS assessment**	**Big Idea written assessment**	**Performance interview**
Properties of sound	Sound quality	Same materials make similar sounds		Identify which objects that make similar sounds (BI_Q1)	
	Sound pitch	Things that affect pitch include: length, tautness, size, etc.	Identify that when a string is pulled tighter, it will make a higher-pitched sound (FOSS_Q2)	Identify that tighter rubber bands make higher-pitch sound (BI_Q2)	Identify a shorter tube can make a higher-pitched sound (PI_2b) Identify looser drum makes lower-pitch sound (PI_3a).
Processes of sound	Sound production	Sound is caused by vibrations		Explain how sound is produced by a guitar (BI_Q3)	Explain that sound is made when xylophone tube vibrates (PI_2a) Explain shorter objects vibrate faster to make higher pitch sound (PI_2c) Explain tighter objects vibrate faster to make higher-pitched sounds (PI_3b)
	Sound transmission	Sound travels through air	Explain how sound travels (FOSS_Q3)		Explain how sound travels (PI_1a) Predict and explain what happened when sound path is blocked by a plastic cup (PI_1b)
		Sound travels better through solids	Identify that sound is louder when travel through solids (FOSS_Q1)		Predict and explain why sound is louder when travels through strings (PI_1c)

Task 2. The interviewer hit a xylophone tube to make sound and students were prompted to explain why sound was produced while being hit. Later, students were asked to pick another xylophone tube among several to produce a higher-pitched sound and explain why a shorter/longer tube would make a higher-pitched sound.

Task 3. Students were presented with three drums made of balloons and plastic cups with balloons stretched over the top of plastic cups. The three drums produced different pitches if tapped on the top. Students first explored the drums but not tapping on top and predicted and then explained which drum would make the highest pitch. After they were allowed to tap on top of the drums, they explained whether they had made the right prediction and why that drum made the highest pitch.

At the end of the sound interview, one measurement question was asked as an indicator of the attainment of the uniformity concept of measurement. This question was to ask students to decide how to measure a table in order to know if it can fit into a certain space. Students can choose among wood blocks, different-sized batteries and broken crayons to measure. To control for the possibility that students had remembered the correct answer from the previous interview for measurement, this question used lego blocks instead of batteries as the correct answer for same-sized units.

3.5 CODING SCHEMES AND DATA ANALYSES

According to the research questions addressed earlier in this paper, data analysis included student learning of measurement and sound concepts, teacher's enactment of the Measurement and Sound modules, and the relations between teacher's enactment and student learning. The analysis of student learning and teachers' enactment was based on teacher or student talk relating to descriptive/factual knowledge (DF), causal explanations (CE) or procedures (Proc). CE requires higher-level thinking including finding data patterns or making generalizations, whereas Proc or DF does not press student thinking beyond making or describing observations. The involvement of DF, CE or Proc in teacher and student talk was the main structure of the coding schemes used in this analysis.

The following analyses were separated into three sections: (1) student talk of measurement and sound concepts after instruction, (2) teacher talk during instruction of the Measurement and Sound modules, and (3) patterns between student talk and teacher talk.

3.5.1 Student talk in interviews and written assessments

Students' learning of Measurement and Sound modules after teachers' instruction was analyzed using their interviews and written assessment data. Students' performance was categorized into: descriptions of the phenomena (DF), causal explanations about scientific phenomena (CE), and procedures about doing science (Proc). Students' talk regarding DF, CE, or Proc was compared among the three classes. Due to the difference of the covered content of Measurement and Sound modules, students' learning was categorized into CE and Proc in the Measurement module and coded into DF and CE in the Sound module. More detailed descriptions of the coding schemes of these two modules were provided in the following section.

3.5.1.1 Coding scheme for student talk—Measurement module

Students' learning of measurement includes not only the ability to execute measurement procedures but also the rationale for employing such procedures. In this coding scheme, two types of learning were coded: the procedures of carrying out the appropriate use of measurement tools (Proc) and the explanations of the reasons for using those procedures (CE).

Descriptive/Factual knowledge about measurement (DF) is not included because the memorization of measurement units (an example of DF in measurement) is not related to the scientific inquiry and thus not considered as one of the learning outcomes in this module. The following learning outcomes of student interviews were coded as either procedures (Proc) or causal explanations (CE) (see Appendix E).

(1) *Procedural understanding (Proc):* students should be able to identify the proper attributes to measure, choose the appropriate tools to measure, and utilize these tools and report results correctly. For example, when trying to separate a fixed amount of food in half, students should be able to decide how to separate a pile of dog food (e.g., count the pieces, weigh it or scoop it) and use proper tools (e.g., balance or measuring cup) to measure equal amount of food. Students were scored by whether they demonstrated the procedures correctly or not.

(2) *Causal explanatory understanding (CE):* students should be able to estimate properly with standard units and to explain the rationale of measurement procedures. For example, when measuring length, students should be able to explain the choice of using same-sized nails (i.e., identical units) instead of long but different-sized nails to measure is because same-sized nails will yield the same results when used by different investigators. Students needed to pick the right tools or use the same tool repeatedly and also to explain correctly about why using the same tool to get the full score. If students picked the right tools but did not provide correct explanations, they were only given partial scores for that item.

The explanation of measurement procedures requires students to make generalizations about measurement and consider the priorities of important criteria about measurement. For example, selecting a tool that produces reliable results is more important than selecting a tool that requires less repetition of measurement procedures. That is, measuring length with same-sized nails to produce reliable results is more valued than measuring with long but different-sized nails. The former produces a smaller-numbered result but may not always conclude with the same answer. Students' explanations of measurement procedures were used as an indicator of students' understanding about measurement in addition to the execution of measurement procedures. The scores of CE and Proc knowledge in the performance interviews and written assessments were compared among the three classes to see whether there was a difference in students' learning of measurement.

The interrater reliability analysis was done by randomly selecting 50% of students' performance interviews and written assessments, which were then coded by another doctoral student. The reliability between two coders was 90% for the performance interview and 98% for the written assessment. Disagreements of the codes between the two coders were resolved through discussions.

3.5.1.2 Coding scheme for student talk—Sound module

In the Sound module, students were involved in the activities of observing the different sounds made by different objects, looking for relationships between sound pitch and object's length, or testing different medium for sound to travel. DF knowledge was considered as phenomena which can be described through observations or from past experiences. For example, the properties of sound (e.g., pitch or volume) can be observed from its high or low pitch or from whether it's loud or quiet. CE knowledge referred to explanations of sound phenomena, such as how sound is produced, how sound travels through different media, or why the length of an object affects sound pitch. Since CE knowledge had to refer to the processes of sound in order to provide explanations, the items related to the process of sound were considered as CE knowledge. Additionally, no procedural knowledge was involved in the learning outcomes of the Sound module, so students' understandings were only categorized into descriptive/factual knowledge (DF) and causal explanation (CE). The followings were examples of DF and CE understandings as identified in the coding scheme for the sound interview (see Appendix F).

(1) *Descriptive/Factual knowledge (DF):* Students should be able to describe sound properties, explain how sound is produced by referring to personal action or materials, identify different pitches of sound, or explain how sound travels by indicating possible paths.

(2) *Causal explanations (CE):* Students should be able to explain sound production or transmission by referring to basic ideas of sound (e.g., sound is caused by vibrations), explain how different pitches of sound are related to the properties of objects (e.g., different lengths affect sound pitch) or talk about the relation between pitch and length (e.g., the shorter tube makes higher-pitched sound).

In this coding scheme, DF understanding involves describing what can be observed, whereas explanatory understandings require students to think about what happened behind the observations, to make a general statement from observations, or to explain phenomena with generalized statements. For example, when being asked why a sound source covered by a cup can be heard outside of the cup, a student may explain that sound can get out from the bottom of the cup. Another student may answer that the sound source is vibrating so the cup is also vibrating. The former explanation described how sound could get out and will be considered as descriptive understanding; the later explanation was based on basic ideas or generalizations of sound and will be coded as explanatory understanding.

Fifty percent of students' interviews and written assessments were randomly selected and coded by another doctoral student to calculate the interrater reliability. The reliability of this coding scheme is 87% for the performance interview and 100% for the written assessment. Differences of the codes were resolved through discussion.

3.5.1.3 Analysis of student talk

Each learning outcome listed in Table 2 and 3 was compared among the three classes using either chi-square analysis or non-parametric Kruskal-Wallis test depended on whether it was categorical or ordinal data. All learning outcomes were clustered into CE or Proc knowledge for the measurement module and into DF or CE knowledge for the sound module. Within each type of knowledge, I summarized significant results to see if students from different classes perform differently or similarly on DF, CE or Proc understanding.

3.5.2 Teacher talk in classroom observations

Teachers' enactment of the curriculum plays an important role on the types of inquiry opportunities provided to students. In order to document teachers' enactment of the modules, field notes from observing each teacher's classes of each module were coded. The taking of the field notes focused on how teachers introduced hands-on activities and made connections to relevant science concepts. Classroom management and topic-irrelevant conversations were not the focus of the study and thus were not included in the analysis.

3.5.2.1 Coding scheme for teacher talk

The unit of analysis was based on what Mehan (1979) called a "Topically Related Set (TRS) of interactional sequences." According to Mehan's (1982) explanations, a unit of TRS could include three-part and extended sequences. A three-part sequence indicates a traditional form of instruction: Initiation, Reply and Evaluation (IRE), which includes an initiation by teacher, followed by student responses and concluded with teacher evaluation. An extended sequence occurs when the immediate reply is not followed by a positive evaluation and thus proceeds with another sequence of teacher initiation, students' replies, until a positive evaluation is provided. Each TRS can be differentiated by teachers' positive evaluation, slowed rhythm or manipulation of educational materials. In this study, the sequence of TRS was also based on the shift of topic or expected student answers.

This coding scheme (see Appendix G) was revised from Newton & Newton (2000) in which they observed teachers' support of causal understanding and descriptive/factual understanding in science lessons. Teacher initiated questions were targeted in this coding scheme to represent the type of scientific thinking that teachers provided in the classroom. Each sequence of TRS was identified together by the author and a doctoral student for each teacher-initiated question. Repeating questions with the same expected student answer were combined as one sequence. Questions with different topics or different student answers were separated into different sequences. For each identified sequence, the content of the sequence was classified to descriptive/factual (DF), causal explanatory (CE) or procedural (Proc) talk.

(1) *Descriptive/Factual talk (DF):* Conversations of DF knowledge include recalling or introducing definitions of vocabulary, summarizing what has been done, or describing what was observed during the activities.

(2) *Causal explanatory talk (CE):* Conversations of CE knowledge involve using generalizations of observations to reach conclusions, make predictions, or apply to new phenomena. Generally speaking, CE talk requires some basic understanding initiated from DF talk, but goes beyond discussions of observations. The generalizations from observations were considered as the main criteria of deciding CE from DF talk.

(3) *Procedural talk (Proc):* Conversations that instruct students to carry out the procedures or explain the procedures of science activities were recorded as procedural talk.

The type of DF, CE or Proc talk that teachers engaged in with their students provided evidence of how teachers enacted the FOSS curriculum and the proportion of different types of scientific thinking encountered in a typical classroom. A doctoral student coded seven out of thirteen of the field notes to calculate the interrater reliability. The reliability was 82% for the coding of teacher talk. Disagreements were resolved through discussion.

3.5.2.2 Analysis of teacher talk

Descriptive analysis such as the average and total occurrences and percentages of DF, CE or Proc talk were conducted to summarize the kind of inquiry opportunities provided in the classroom. The average proportion of DF, CE or Proc talk represented the focus of teachers' instruction and the types of understanding emphasized through question-asking interaction. Higher percentages of DF talk among the three teachers indicated more emphasis on DF knowledge, lower percentages of CE talk showed less focus on CE knowledge, and so on.

3.5.3 The relation between student talk and teacher talk

The third section of data analysis was to find out the relation between teachers' enactment of FOSS (teacher talk) and students' learning of science concepts (student talk) regarding DF, CE or Proc. Table 4 summarized DF, CE and Proc understanding in both teacher talk and student talk. The results of students' DF, CE and Proc understanding provided information on how well students learned about describing, explaining and doing scientific activities. The percentages of teachers' DF, CE and Proc talk represented the focus of instruction. For those significant differences between classes in student talk, I compared the percentages of the corresponding type of teacher talk and examine if similar patterns occur. For example, if students from class A showed higher performance on CE understanding than students from class B, then the percentages of teachers' CE talk in class A would be compared with class B to see if A is higher than B. If the analysis of student talk and teacher talk shows the same pattern, it indicates that teachers' enactment and student performance may be positively related.

If a significant result in student talk is not found in teacher talk, then a direct relation between teacher enactment and student learning would not be found in these data.

Table 4. Summary table of the coding schemes on students' learning and teachers' enactment

	Coding scheme of teacher talk	Coding scheme of student talk	
Codes		Measurement	Sound
Descriptive/ Factual (DF)	▪ Describe phenomena ▪ Recall facts/vocabulary ▪ Recall observations	n/a	▪ Describe observable facts about sound properties/production/propagation
Causal Explanatory (CE)	▪ Aim to find out patterns of data ▪ Aim to find a generalization of patterns	▪ Make reasonable estimations using standard units ▪ Explain the rationale of doing particular procedures	▪ Explain sound production or transmission based on big ideas (i.e., sound is caused by vibrations)
Procedural (Proc)	▪ Instruct about carrying out the activities	▪ Execute measurement procedures	n/a

4.0 RESULTS

This study was aimed at exploring the impact of a professional development workshop on students' learning and teachers' enactment, and documenting the implementation of hands-on science in actual classrooms. Students' and teachers' talk about descriptive/factual (DF), causal explanatory (CE), or procedural (Proc) knowledge was analyzed to see if there is any difference between the classes of the target teacher and the comparison teachers. Moreover, the patterns between students' DF/CE/Proc knowledge and teachers' talk in the classroom were explored to describe the resulting differences among the three classes. This chapter started by comparing students' learning in the measurement and the sound units, followed by summarizing different types of teachers' talk in the classes, and concluded by the exploration of the relation between student learning and teacher talk.

4.1 THE IMPACT ON STUDENTS' LEARNING

The first part of the results focused on the impact on students' understanding of science knowledge, including their descriptive knowledge (DF), explanatory knowledge (CE) and procedural knowledge (Proc). The data on students' learning outcomes were collected from their written assessments and performance interviews and were categorized into DF, CE or Proc knowledge based on the identified big ideas within each FOSS modules. Students' learning outcomes were compared among the students of the target teacher and the two comparison teachers. Due to the fact that the learning outcomes from the written assessments and performance interviews were either coded as ordinal scales or categorical data, the analyses employed non-parametric Kruskal-Wallis test for ordinal-scale items (i.e., the degree of sophistication on student explanations) and chi-square test for percentage-of-correct-responses. If significant differences among the three classes were found, the comparison between each pair of classes (post hoc test) was then conducted using Mann Whitney U test.

4.1.1 Students' learning of measurement

Students' understanding of measurement was separated into procedural (Proc) knowledge, such as carrying out the measurement procedures, and causal explanatory (CE) knowledge, such as making estimations with standard units, or using identical units to measure. According to the identified big ideas and learning goals, which were listed in Table 2, each learning outcome was categorized into measurement procedures (Proc knowledge) or measurement concepts (CE knowledge). The presentation of the results started with students' Proc knowledge, and then followed by their CE knowledge.

4.1.1.1 Procedural knowledge: Learning outcomes of measurement procedures

The following section combined the results from both written assessment and performance interviews which students took after the completion of the measurement module. The learning outcomes which related to measurement procedures were grouped into (1) to identify appropriate property of interest for measurement, (2) to use measurement tools appropriately, and (3) to interpret measurement results appropriately. The detailed results were shown in Table 5.

Table 5. Percentages of correct response of measurement procedures

Measurement Procedures	Sarah (n=13)	Clark (n=10)	Margaret (n=15)	Chi-sq. (df=2)	Prob.
-Identify property of interest:					
Measure the space in between (WA†: BI*_2a)	92%	83%	80%	.719	.698
Compare the size of books (PI†: 1d)	46%	58%	36%	1.121	.571
-Select appropriate tools:					
Choose thermometer to measure temperature (WA: FOSS_4)	92%	82%	73%	1.703	.427
Choose rulers to measure length (WA:BI_2c)	92%	83%	67%	2.713	.258
Choose rulers to measure length (WA:BI_3b)	92%	92%	73%	2.564	.277
Choose rulers to measure length (PI: 1c,1d)	100%	100%	100%		
Use tools to separate dog food (PI: 2a): without prompting with prompting	0% 69%	17% 68%	20% 71%	2.819 2.906	.244 .234
Use tools to separate water (PI: 2b): without prompting with prompting	77% 85%	75% 83%	93% 100%	1.747 2.507	.417 .285
-Place materials appropriately:					
Draw marbles to make a scale balanced (WA: BI_1b)	67%	92%	80%	2.304	.316
Measure length of a book (PI:1c)	100%	100%	100%		
Separate dog food and water in half (PI: 2a)	100%	100%	100%		

-Report results properly:					
Report results off a balance scale (WA:FOSS_14)	15%	17%	7%	.743	.690
Report results off a thermometer (WA:FOSS_24a)	100%	91%	100%	2.612	.271
Report results off a thermometer (WA:FOSS_24b)	92%	80%	100%	4.422	.352
Report results off a graduated cylinder (WA:FOSS_26)	77%	91%	87%	.974	.614
Predict results of a scale (WA:BI_1a)	46%	42%	73%	3.310	.191
Balance a scale (PI:2a)	100%	100%	100%		
Read a measuring cup (PI:2b)	100%	100%	100%		

†WA: written assessment; *BI: Big Idea assessment items; †PI: Performance Interview items

(A) Identify property of interest.

In the written assessment and performance interviews, students have to identify the width in between two pieces of furniture as the property of interest to determine whether a desk can be placed in the bedroom, identify length and/or width to measure when comparing the size of two books, and identify weight or volume to measure while being asked to separate dry dog food in halves.

The results on student performance were similar between Sarah's students and students of the other two comparison teachers. All of the students were able to identify properties of interest according to different measurement purposes. However, when the property of interest consisted of more than one variable, students were not always clear about what to measure. For example, one possible way to measure about the size of a book can be the area of the book cover. Students were able to identify at least one variable (length or width) to measure but only 47% of the students mentioned to use more than one variable (length, width or height) to represent the book size.

The consideration of multiple variables at a time requires more sophisticated thinking in child's development. Problem-solving tasks which involved multiple variables also showed that children did better on predicting or generalizing outcomes regulated by one factor, but performed poorly when dealing with outcomes controlled by multiple factors (Schauble, 1990; Thornton, 1995). Younger children tend to focus on the influence of one factor only, and later on they may notice the outcomes being integrated by other factors as well, but they will not be able to predict how the outcome will turn out from knowing the relating factors until even later (Wilkening & Huber, 2002). Using multiple aspects to represent one property of objects may not seem as complicated as predicting the outcome variable by manipulating the relevant factors, the measurement task in this study was similar in the sense that children have to understand the outcome variable (the property of interest) is not determined by simply length or width, but the combination of both.

(B) Use measurement tools appropriately

Several learning outcomes were related to the use of measurement tools, including selecting an appropriate tool according to the property of interest, placing materials/objects appropriately, and reporting the results accordingly. In the written assessment, students were asked to identify what tool to choose to measure length or temperature, to predict how to make a scale balance, to read the results from a balance scale, a graduated cylinder, a ruler and a thermometer, and to predict what a scale would look like with unequal weights. In the performance interview, students were asked similar things but they were required to demonstrate how to use the measurement tools and report the results.

Students performed similarly on using the measurement tools across the three classes. If they were asked to pick a tool, students were able to select appropriate tools to measure according to different properties (e.g., length, weight, volume or temperature), but they did not always remember to adopt tools to complete the tasks. In one of the performance tasks which students were asked to separate dry dog food in half. Most students started the task by counting pieces of dog food and did not think of using any tool even though a balance scale and a beaker was right in front of them. It seemed that students at this age still prefer counting strategy of measurement and did not fully appreciate the expedience and convenience of adopting measurement tools.

Moreover, although students may possess the skills to carry out measurement procedures, written assessments did not always reflect students' ability to use these skills. The results on students' performance interviews showed higher percentages of correct responses than the written assessments regarding the same learning outcomes. For example, all of the students demonstrated using a balance scale or a graduated cylinder to measure food or water during their performance interviews, but only 67% to 92% of students answered correctly on their written assessments about how to use tools. These results suggested that there existed comprehension problems among third-grade students regarding the reading and understanding of written texts. Performance tasks may be a more accurate way to assess students' procedural knowledge than written assessments.

(C) Interpret results properly

Students generally reported the results in a similar way by reading the number off a measurement tool. Such behavior may lead to some problems while interpreting results of fractions or from a broken ruler. These measurement behaviors occurred commonly among the three classes and were noted below. One was the results involving fractions. Even though students at this age seemed to have some ideas that something can be just half of a unit, they would report the results as “one and a half” unit, using a standard tool or a nonstandard tool. For example, if an object is 3 ½ cm or 3 and a half crayons long, many students would conclude that its length is 4 ½ (cm or crayons). Students’ tendency to report fractional results as one-and-a-(fraction) may take more instructional activities to fully comprehend, and thus should be noted in the curriculum before students encountered fractions during measurement activities.

Another inappropriate reporting of results was that students seemed to just follow the measurement procedure to read any number off a tool without considering the origin of the tool. It occurred in one of the FOSS assessment item when two pictures consisted of the same nail along with a ruler were shown to students. The only difference was that one nail was aligned to the origin point and the other nail was placed along a broken ruler. Students just read off the number on a ruler while reporting the results. Only about 7% to 17% of students knew that the numbers on a broken ruler does not represent the actual length of an object and proposed other strategies of measuring the length. Such misconceptions about measuring procedures may need to be further explored during the measurement activities.

Students revealed some problems with reporting the results, mostly from length measurement. But overall, students’ performance of procedural knowledge was consistent among the three classes. That is, they were able to identify the appropriate property of interest, use measurement tools, and report measurement results (if there was no fraction or broken ruler involved).

4.1.1.2 CE knowledge: Learning outcomes of measurement concepts

Students’ learning of measurement concepts was categorized as CE knowledge. Although students did similarly well on their measurement skills, they performed differently on their learning of measurement concepts. Two measurement concepts were included in the study: (1) making appropriate estimations with standard units, and (2) using identical units to measure. The learning outcomes that showed differences among the three classes were all related to making reasonable estimations (see Table 6). The concept of uniformity, however, did not reveal any significant differences in students' performance among the three classes. The comparison between each classes (post hoc test) were provided in Appendix H.

Table 6. Percentages of correct response of measurement concepts.

Measurement Concepts (CE knowledge)	Sarah (n=13)	Clark (n=10)	Margaret (n=15)	Chi-sq. (df=2)	Prob.
-Making estimations with standard units					
Estimate length of a piece of paper (WA:FOSS_3)	92%	25%	73%	10.433	.005*
Estimate height of an adult (WA:FOSS_15)	46%	33%	40%	.445	.801
Estimate mass of a piece of paper (WA:FOSS_11)	15%	25%	60%	6.231	.044*
Estimate length of a book (PI:1b)	85%	45%	93%	8.845	.012*
-Using identical units to measure					
Identify measurement tools with identical units (WA:BI_3)	38%	8%	13%	4.207	.122
Identify measurement tools with identical units (PI:1e)	36%	50%	73%	3.705	.157

(A) Make appropriate estimations with standard units

On the written assessment, three out of four items relating to making estimations showed significant differences. For the two items about making length estimations, more students in Sarah's (92%) and Margaret's (73%) classes made correct estimation about the length of a piece of paper than in Clark's (25%) (χ^2= 10.433, p<.01) in FOSS item #3. Another item (FOSS #15), which also related to length estimation, did not show any difference among the three classes. This item asked students to estimate an adult's height. Since this item was presented with choices in different units (e.g. cm, mm, km, m), the item was targeted toward the distinctions in standard units rather than estimating a reasonable value for height. Students from the three classes performed similarly on this item.

For the item about estimating the mass of a piece of paper (FOSS #11), 60% of Margaret's students answered correctly whereas only 15% of Sarah's and 25% of Clark's students did (χ^2= 6.231, p<.05). Margaret's students made a better estimation on both length and weight, Sarah's students had a better grasp of estimating length than weight, and Clark's students did not make good estimations of either length or weight. The estimation of length measurement indicated that Clark's students might not have a clear mental representation of any of the standard units.

On the performance interview, students' learning outcomes also showed significant differences between Clark's students and students of the other two teachers. As shown in Table 6, most students of Sarah's (85%) and Margaret's (93%) classes made reasonable estimation of length (within 15cm of the actual length of a book), but only 45% of Clark's students made reasonable estimation (χ^2= 8.845, p=.012). Clark's students made wild guesses such as 100cm or 6cm for a book that is 35cm long. It seems that they did not have a clear sense about the actual length represented by centimeters.

Moreover, some students used strategies to help them estimate, such as moving their pinkie (estimated as 1cm) through the book or used their fingers to make a small range (estimated as 10cm) then go through the book. Such strategies seemed to help students come up with a better guess, and only students of Sarah's and Margaret's students came up with estimating strategies during the performance interviews.

(B) Use identical units to measure

Another measurement concept being tested was the concept of uniformity. On the written assessment, students were asked to identify a better way to measure a jump rope by choosing from same-length pencils, different-length crayons and different-length blocks. The correct response was to choose pencils and justify that pencils were the same size. Based on students' choice and justification all together, Sarah's students (38%) were more likely to provide justifications consistent with the uniformity concept than Margaret's (13%) and Clark's (8%) students, though these differences were not statistically significant. Some students chose same-sized pencils (the correct answer) but provided justifications which did not show any understanding of the uniformity concept, such as "it's the longest" or "it's the same as jump rope." It implied that some students did not fully understand the priority of uniformity when choosing measurement tools.

During the performance interview, the uniformity concept was tested by asking students to choose a non-standard tool to measure the length of a book with either same-sized batteries, different-sized screws or broken crayons. The results did not show any significant difference among the three classes either (χ^2=3.705, p=.157). Margaret's students (73%) showed much higher percentages of choosing same-sized batteries than Sarah's (36%) and Clark's (50%) students. There were two possible responses which were considered as correct. Students may line up batteries which are all the same size or repeat using the same tool (e.g., a screw or crayon). Both would show students' understanding of using identical units.

A few students seemed to show better understanding about using same-sized tools and explained clearly that "*if one [is] bigger than the other, you'll end up with different answers.*" Other students who did not choose identical-unit tools were eager to choose a longer tool for faster measurement or to find a tool to be flush with the measured object. These students may still be confused that the primary goal of measurement was not to expedient measurement or to fit end-to-end for measurement tools, but to be able to repeat measurement outcomes reliably across different time and space.

In summary, students performed differently on the items of measurement concepts. Higher percentages of Sarah's and Margaret's students were able to make reasonable estimations of length measurement compared to Clark's students on both the written assessment and the performance assessment. However, students' understanding of the uniformity concept did not show any difference among the three classes. It is likely that some measurement concepts take more in-depth activities for students to fully comprehend, and the hands-on activities which stressed mostly the practice of measurement skills do not automatically nor easily lead to the comprehension of basic notions of measurement.

4.1.2 Students' learning of sound

Students' learning outcomes in the sound module were separated into two types of content knowledge, the properties of sound and the processes of sound. The knowledge about the properties of sound may include describing different attributes of sound, such as low or high pitches, or distinguishing the sounds made from different materials. This type of knowledge was mostly about making observations of sound properties and was categorized as the descriptive (DF) knowledge of sound in the analysis. The knowledge about the processes of sound requires the explanation of sound production or sound transmission which needs to go beyond descriptions or observations of sound phenomena. This type of knowledge usually involved with the explanation of sound processes and was categorized as causal explanatory (CE) knowledge in the analysis. No procedural knowledge was identified as learning outcomes in the sound module.

Based on the big ideas and learning outcomes targeted in the sound module, students' data from written assessments and performance interviews were categorized into: (1) *properties* of sound: such as identifying sound pitches, or identifying factors affecting sound pitches, and

(2) *processes* of sound: such as explaining how sound is produced or how sound travels from one place to another. The number of participating students was different in the written assessment and the performance interview[3] and was summarized in Table 7.

Table 7. Participating students in the sound module.

	Sarah	Clark	Margaret
Written Assessment	5	11	15
Performance Interview	13	7	15

4.1.2.1 Descriptive/Factual (DF) knowledge: Properties of sound

Understanding about the properties of sound was categorized as the descriptive (DF) knowledge about sound and was focused on the factors which affect the property of sound. The learning outcomes included (1) to identify materials which made similar sounds, (2) to identify the relations between length and pitch, and (3) to identify the relations between tautness and pitch. The results of students' learning outcomes on the property of sound were summarized in Table 8. The comparison between each pair of classes (post hoc test) was done by Mann Whitney U test and the results were provided in Appendix I.

Table 8. Results of students' learning of sound properties.

	BIG IDEA	**Sarah**	**Clark**	**Margaret**	**Chi-sq. (df=2)**	**Prob.**
(A)	Same materials make similar sounds (WA†:BI_Q1)	100%	18%	50%	9.363	.009**
(B)	Identifying length as a factor of sound pitch of tubes (PI‡:#2)	100%	67%	100%	9.917	.007**
(C)	Tighter rubber band makes a higher sound (WA:BI_Q2)	80%	9%	43%	7.876	.019*
	Tighter string makes a higher sound (WA:FOSS_Q2)	20%	18%	43%	2.072	.355
	Identifying tautness as a factor to sound pitch of drums (PI:#3)	31%	14%	60%	4.890	.087

†WA: written assessment; participating students: Sarah (n=5), Margaret (n=15), Clark (n=11).
‡PI: performance interview; participating students: Sarah (n=13), Margaret (n=15), Clark (n=7)

(A) Same materials made similar sounds

This learning outcome was only assessed in the written assessment. Students from the three classes showed a significant difference in this item (WA:BI_Q1). Sarah's students performed best among the three classes (χ^2=9.018, p=.011). All of Sarah's students correctly identified

[3] Sarah only had time to administer the written assessment to one of her classes, and only five students participated in the written assessments. In Clark's class, the substitute teacher kept putting off the interview schedule till the very last class of the semester. There were not enough time to interview all of Clark's students and only seven were interviewed.

objects of the same material as those which make similar sounds, whereas 47% of Margaret's students and only 18% of Clark's students showed such understanding. Most of Clark's students (64%) and 40% of Margaret's students chose the objects with similar shape as the ones which make similar sound. It indicated that they did not consider material to be more important to the property of sound than shape.

(B) Relation between sound pitch and property of objects—length

Students were also assessed on whether they can identify the relationship between sound pitch and the property of objects. One of the tested properties of object was its length. In the performance interviews, students were asked to first predict, among the xylophone tubes with different lengths, which object would make the highest pitch sound. Students' performance showed difference in identifying length as a factor related to sound pitch. All of Sarah's and Margaret's students correctly identified that different lengths would make different pitches of sound, but only 67% of Clark's students clearly identified length as the factor ($\chi^2=9.917$, $p=.007$). Thirty-three percent of Clark's students mentioned several factors which may or may not include length, but they also mentioned other factors irrelevant to pitch. These students were considered as being unsure about the effect of length on sound pitch.

(C) Relation between sound pitch and property of objects—tautness

In addition to length, students were also asked to identify tautness as a factor to sound pitch. Different objects were used to test this learning outcome, including rubber band and drums. In the written assessment, pictures of one thinner/tighter rubber bands and one thicker/looser rubber band were shown to students and students had to identify which one would make a higher pitch (WA:BI_Q2). Sarah's students performed better than the comparison students on this item. Eighty percent of Sarah's students chose the tighter rubber band (the correct answer), which was higher than 40% of Margaret's and 9% of Clark's students who answered correctly ($\chi^2=7.557$, $p=.023$).

Looking at the distribution of the rubber band item, many of Clark's students (64%) chose an incorrect answer that the looser rubber band would sound louder, which indicated that they did not understand which factors affect volume and which factors affect pitch or they were confused about the vocabulary of loud and high-pitched sound. Thirty-three percent of

Margaret's students chose that the looser rubber band would make a higher-pitched sound, which indicated that they had some understanding of what affected sound pitch but were confused about the relationship between pitch and tightness.

Another written assessment item assessed a similar concept but the results were different from the rubber band item for Sarah's students. This item (WA:FOSS_Q4) asked how might the pitch change if a string tied on a doorknob was tightened. Twenty percent of Sarah's students selected the correct answer which said "the pitch will be higher as you tighten the string." The correct percentages of Margaret's and Clark's students remained similar to their respective performance in the previous item. The inconsistent results of Sarah's students can be due to students' partial understanding of the tautness of the string or their confusion about the doorknob activity. It might indicate that Sarah's students did not fully understand that the tautness of the string was the same as pulling the rubber band tighter or students may have been confused about the doorknob activity from their science class.

On the performance interview, although the property of the tautness of drums seemed similar to the tightness of strings, it was more challenging for students to identify tautness as a factor to the pitch of drums. Smaller percentages of the students (31% of Sarah's students, 60% of Margaret's students and 17% of Clark's students) identified the tautness of drums compared to the correct response regarding the tightness of strings.

In summary, the results generally showed that Sarah's and Margaret's students tended to outperform Clark's students on relating concepts about sound properties (DF knowledge). However, the items or tasks related to the tautness of objects did not all yield consistent results within each class.

4.1.2.2 Causal explanatory (CE) knowledge: Processes of sound

For the processes of sound, the descriptions of sound production and transmission are non-observable from the performance tasks or the assessment items. In those questions, students had to apply their understanding about sound to explain how sound is made or how sound travels. Therefore, the ideas about the processes of sound in their responses were categorized as students' understanding of CE knowledge in the sound module.

The analysis of students' explanations contained both percentages of correct responses and the mean of the average scores assigned to each student. The results were shown in Table 9 and divided into four sections: (A) sound production: explain that sound is caused by vibrations; (B) sound production: explain the relation between sound pitch and property of objects; (C) sound transmission: explain that sound travels through air; and (D) sound transmission: explain that sound travels through strings.

Table 9. Results of students' learning about sound processes.

	BIG IDEA	Sarah	Clark	Margaret	KW test (χ 2)	Prob.
(A) Sound production: vibrations	Explain how guitar made sounds. (WA†:BI_Q3)	Mean:1.00 (SD: 0)	0.82 (0.405)	1.36 (0.497)	8.033	.018*
	Explain how xylophone tubes made sound. (PI‡:#2)	1.23 (0.599)	1.00 (0.632)	1.80 (0.414)	10.107	.006**
(B) Sound production: pitch	Explain why shorter tubes made higher pitches. (PI:#2)	0.23 (0.438)	0 (0)	0.60 (0.507)	7.892	.019*
	Explain why tighter drums made higher pitches (PI:#3)	0 (0)	0 (0)	0.133 (0.352)	2.747	.253
(C) Sound transmission: thru air	Explain how sound travels (WA:FOSS_Q3)	1.40 (0.894)	0.45 (0.522)	1.57 (0.646)	12.286	.002*
	Explain how sound travels to your ears (PI:#1)	2.38 (0.768)	1.57 (0.976)	2.33 (0.9)	3.799	.150
	Explain how sound travels when covered with a cup (PI:#1)	2.23 (0.725)	1.86 (0.69)	1.93 (0.799)	1.566	.457
(D) Sound transmission: thru solids	Sound is louder when travels through solids (FOSS: Q1)@	40%	9%	57%	6.154	.046*
	Comparing sound travels with or without a string (PI:#1)	3.46 (1.506)	3.14 (1.864)	3.13 (1.407)	0.408	.816

†WA: written assessment; participating students: Sarah (n=5), Margaret (n=15), Clark (n=11).

‡PI: performance interview; participating students: Sarah (n=13), Margaret (n=15), Clark (n=7)

@ : this item is a multiple choice item, so the analysis was based on Pearson chi-square to compare the percentage of correct response among the three classes.

(A) Sound production: Explain that sound is caused by vibration

The targeted idea about sound production was that sound is caused by vibrations. Different objects were used while testing students' explanations of sound production. When showing students a picture of a person playing a guitar and asking why guitars make sounds, students tended to refer to personal actions (e.g., "because you plunk it", or "because someone plays it") instead of the vibrations of the strings. Responses which referred to the vibrations of the strings were given full credit, and those referred to personal actions were given half credits since it did not answer how sound is made. Margaret's students did significantly better than Clark's students (χ^2=8.033, p=.018), but Sarah's students did not differ from either Margaret's students or Clark's students. Margaret's students (33%) correctly mentioned that sound is caused by vibrations to explain how sound is produced from a guitar (e.g., "the string[s] vibrate and make sounds" or "the vibrations in the strings"). However, all of Sarah's students and 82% of Clark's students mentioned that sound is caused by personal actions, instead of referring to the vibrations of the strings. It was possible that this short answer item did not correctly reflect students' understanding since students might misunderstand the meaning of the question by referring to personal actions.

In the performance interview, students were further prompted to explain how sound is produced. During the interview, one xylophone was presented and students were asked to explain why a xylophone tube made a sound. More students (33% more in total) talked about the vibration of the tube made the sound in the interview than in the written assessment. Nevertheless, many students still referred to personal action (e.g., "because you hit it") as the cause of sound. The prompts by interviewers (e.g., "what happened to the tube that made a sound") did not make much difference. Margaret's students had significantly higher scores than Sarah's and Clark's students (χ^2=10.107, p=.006).

(B) Sound production: Explain patterns between property of objects and sound pitch

In addition to explaining how sound was made, students were also asked to explain why shorter xylophone tubes made a lower sound (or a higher sound depending on their answers). Although students were able to identify the relationship between length and sound pitch, most of the students did not provide an explanation for the patterns between pitch and length. The common answer of the students was to restate the correlation between pitch and length by saying "*because it's shorter*" without any attempt to include big ideas in their explanations (e.g., *because it's shorter so it can vibrate faster*). Some students would come up with their own mechanism to explain. For example, some students said "*it's shorter and the sound doesn't have to take much time to come out, so it [makes a] higher [pitch]*." Although such explanations were not correct, students were given partial credits due to their attempts to go beyond observations and provide a mechanism as explanations.

No student explained correctly that a shorter tube will vibrate faster than a longer tube to make a higher pitch. Margaret's students scored significantly higher than Clark's students on providing explanations of sound production (χ^2=7.024, p=.030). The scores of Sarah's students did not differ from Margaret's or Clark's students. Sixty percent of Margaret's students provided their own explanations on why shorter/longer tubes make higher/lower pitches. Twenty-three percent of Sarah's students also provided their own mechanism, but none of Clark's students came up with their own explanations to the relations between pitch and length of objects.

For the relationship between tautness of drums and sound pitch, it was more difficult than the xylophone tube task. Students were not sure the reason that some drums made a higher pitch was due to their tautness at the first place. Therefore, students also had problems explaining correctly that the air inside the tighter drums vibrates faster so the pitch is higher or even providing their own mechanism. Students did not differ from each other on providing explanations about the relation between sound pitch and tautness (χ^2=4.695, p=0.096).

In summary, results of sound production suggested that more of Margaret's students were able to use big ideas or general principles (e.g., sound is caused by vibrations) to explain why something made sounds whereas Sarah's and Clark's students mostly talked about sound being made by human actions. Most students were not aware of the difference between describing the patterns and explaining the patterns with big ideas or general principles. Margaret's students were significantly better than Clark's students at providing explanations for scientific phenomena, regardless the correctness of the explanations. Some of Sarah's students were able to provide their own mechanism to explain the patterns of sound pitch and length, but not as many as Margaret's students.

(C) Sound transmission: Explain that sound travels through air

The medium that sound travels through can be gases, liquids or solids. In the written assessment and performance interviews, only two types of medium (i.e., air and string) were assessed. The scores among the students of the three teachers were different in one of the written assessment items which asked students to explain how sound travels from one place to another (χ^2=12.286, p=.002). None of Clark's students referred to air while explaining sound transmission. Forty-five percent of their answers mentioned that sound was loud so that it traveled, whereas 60% of Sarah's and 60% of Margaret's students talked about air in their explanations.

However, such results were not statistically significant in the performance interview. In the performance task, students had to explain how sound travels to their ears when the sound source (a cell phone) was (1) placed on the table and (2) covered by a plastic cup. Higher scores were given when students used big ideas to explain, such as "*the vibrations from the phone come to* *my ear*" or when students mentioned the medium through which sound travels, such as "*the sound comes to my ear because of the air*." While students used big ideas to explain scientific phenomena, it showed that they were not just describing what they saw but also attempted to use general principles to account for different phenomena. Lower scores on this task were the responses which did not show any understanding about how sound travels, such as "*because the sound is loud so I can hear it*" or "*the sound came to my ears*."

Although no significant difference was found by comparing the average scores of the three classes, Clark's students scored significantly lower (in terms of percentages) than those in other two classes (χ^2=6.908, p=.032). The descriptions or explanations of Clark's students were usually very brief and did not show much understanding about the processes of sound. Seventy-one percent (five out of seven) of Clark's students scored the lowest points by referring to "*because the sound is loud*" or "*because I can hear it*." Many of the responses of Sarah's and Margaret's students were more sophisticated than Clark's students. Many of their responses showed that they understood that sound spreads around the room, and the vibration of sound comes to your ear. Some students also showed that they had acquired some vocabulary about sound, such as sound waves or molecules, in their explanations, but they did not always use the vocabulary in a proper way.

Students' explanations of sound also revealed several preconceptions about sound, such as seeing sound being made of matter that can only escapes from cracks of objects (Driver, et al., 1994; Mazens & Lautrey, 2003, viewing sound as a substance that was hidden inside an object and released while the object was hit (Driver, et al., 1994), or mentioning sound travels in one direction (Mazens & Lautrey, 2003). Children usually were able to predict some sound phenomena, but they performed poorly on referring to sound processes to explain sound phenomena. Moreover, students' preconception of seeing sound as a substance is very common in students' explanations. Seeing sound as a substance is the most resilient model that children tend to carry for the longest time among other misconceptions (Eshach & Schwartz, 2006; Lautrey & Mazens, 2004; Mazens & Lautrey, 2003; Reiner et al., 2000). The differences between the properties of matter and energy can be properly introduced in this module if students are expected to understand the sound phenomena they observed rather than to simply memorize the factors affecting sound pitch.

(D) Sound transmission: Explain that sound travels through string

The results of one multiple choice item from the written assessment (FOSS_Q1: sound is louder when travels through solid materials) showed that the percentage of correct responses of Clark's students was lower than Sarah's and Margaret's students (χ^2=6.154, p=.046). Only 9% of Clark's students were able to identify that sound would be louder in solids whereas 40% of Sarah's or 57% of Margaret's students selected the correct answer. Many (45%) of Clark's students chose that sound will be the same in "solids", "liquids" and "air" (the exact terms used in the FOSS item). This indicated that those students did not understand about sound traveling differently with different mediums.

Referring to air while explaining sound transmission seems to be more difficult than when sound travels through solids since air is not visible and can easily be ignored in students' responses. In the performance interview showing two cups attached together with a string, students from the three classes performed equally well on their scores explaining how sound travels (χ^2=0.408, p=.816). They seemed to have some understanding that sound can travel through different mediums and that sound is louder when travels through strings. Some students were able to provide pretty good explanations that "*sound is louder through the strings because it won't spread out like in the air,*" or "*the vibrations of the string make the sound louder*."

In summary, the performance of the students from the three classes showed more differences in their scores on sound production than sound transmission. Margaret's students seemed to be better at providing explanations of sound phenomena than Clark's students. The performance of Sarah's students was in between Margaret's and Clark's students and did not significantly differ from either of those classes.

Students were able to provide sound trajectory while explaining sound transmission, especially when the medium was solid. The mechanism of the production of sound was more ambiguous in students' explanations. The idea that sound is caused by vibrations did not usually occur in their explanations. Students were more likely to attribute the cause of sound to personal action (e.g., someone hits it). Although it is not wrong to explain that sound was caused by personal actions, it is less sophisticated than incorporating ideas about the vibrations of objects into their explanations.

4.1.2.3 Brief summary of students' learning about sound

Comparing students' learning on the sound unit among the three classes, students of Sarah's and Margaret's classes performed similarly and Clark's students differed from the other two classes on a couple of learning outcomes. On sound properties, Clark's students differed from Sarah's students on three targeted ideas: (A) same materials make similar sounds, (B) length affects sound pitch, and (C) tautness affects sound pitch, noting that only some of the items of these targets ideas were statistically different

As for sound processes, Margaret's students usually scored the highest among the three classes. Sarah's students scored slightly lower, but did not differ from Margaret's students. Clark's students were different from Margaret's students on (A) using vibrations to explain sound production, (B) explaining why shorter tubes made higher pitches, (C) explaining how sound travels, and (D) identifying that sound is louder when travels through solids. When talking about the production of sound, Margaret's students were more likely to employ big ideas (e.g., sound is caused by vibrations) in their talk whereas Clark's students only referred to personal actions (e.g., because you hit it) in their explanations. Sarah's students were combined with both explanations and did not differ from the students of either class.

From students' explanations of sound transmission, Sarah's and Margaret's students were more capable of providing elaborated answers about how sound travels or referred to air while explaining it. The talk of Clark's students was rather simple and brief, without much reference to related vocabulary about sound (e.g., sound waves or sound echoes). Thus, Sarah's and Margaret's students may have better understanding of sound transmission than Clark's students.

Moreover, throughout the performance interviews, Sarah's students referred to the related big ideas more often (although this difference was not statistically significant) than the other teachers across the items about the transmission of sound (Figure 5). Most of Clark's students talked about what they observed but did not include big ideas into their answers. However, due to the limitation that the sample size was small and the variations were relatively large on these items, the differences were not significant in this study.

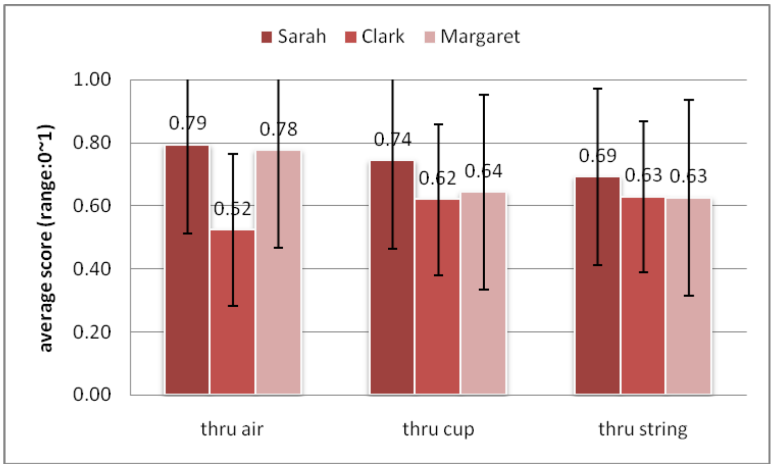


Figure 5. Standardized score of students' talk regarding sound transmission from the performance interview

4.2 THE IMPACT ON TEACHERS' TALK

Teachers' talk about DF/CE/Proc knowledge during their instruction may help explain why students performed differently in their learning outcomes. The following results summarized the percentages of inquiry opportunities the teachers provided for students. The averages of DF, CE and Proc talk were listed to compare the classes of the three teachers. For the measurement module, three field notes from each teacher were collected, and for the sound module, three field notes were collected for Sarah's class, one field note was collected for Margaret's class and no field notes were collected for Clark's class.

4.2.1 Comparison of teacher talk

The analysis of teacher talk used teacher-initiated questions as each unit and coded the types of knowledge that these questions are intended to ask. The author and a doctoral student discussed and determined together what counted as one unit based on different Topic Related Sets (Mehan, 1982) prior to the coding of teachers' talk. Then the predefined units were coded into DF, CE, Proc knowledge or Others. The average and percentages of teachers' talk of DF/CE/Proc knowledge were presented in Table 10.

Table 10. Frequencies and percentages of DF/CE/Proc talk during the instruction of measurement module.

	Sarah	Clark	Margaret
DF	55% (0.212)	32% (0.199)	51% (0.226)
CE	0% (0)	2% (0.027)	1% (0.017)
Proc	26% (0.195)	44% (0.279)	41% (0.171)
Other	19% (0.099)	23% (0.134)	7% (0.069)
Total units	21.33	31.33	28.33

4.2.1.1 Teacher questions on the measurement module

Teachers' questions mostly fell into the categories of DF and Proc knowledge. Comparing the three classes, the highest percentage of teacher questions in Sarah's and Margaret's classes was about DF knowledge (55% and 51% respectively), whereas the highest percentage in Clark's class was about Proc knowledge (44%). Margaret and Clark had similar percentages on Proc talk which indicated that they emphasized similar proportions of procedure-related knowledge in their questions. Margaret had similar percentages of DF talk as Sarah (51% for Margaret and 55% for Sarah), implying that Margaret and Sarah both paid attention to initiating talk about DF knowledge with their students. The percentage of CE knowledge was only minimal across three classes.

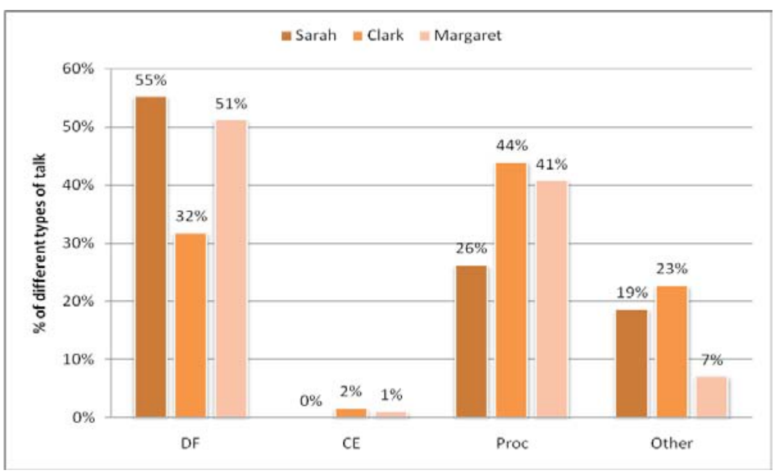


Figure 6. Comparison of different types of teacher talk among the three teachers.

It should be noted that teacher talk of Proc knowledge might be underestimated because teachers usually spent time giving out directions about hands-on activities rather than asking questions about procedures. But the questions that teacher initiated could reflect the type of knowledge that teacher valued and wanted students to understand.

Although Proc talk was necessary to instruct students how to set up the hands-on activities, emphasis on DF or CE talk was also important to provide opportunities to engage in deeper scientific thinking, such as exploring the patterns of their observations, finding explanations about the patterns, or connecting their observations with their past experiences. Without teachers' guidance through DF or CE talk, students were less likely to engage in any meaningful scientific inquiries and may end up practicing skills of manipulating materials.

One thing to be noted is the quality of DF and CE talk. In Clark's class, although 32% of his talk was DF talk and 2% was CE talk, his questions were not about measurement concepts. The questions either focused on memorizing or repeating factual knowledge (e.g., asking "*what's our standard for measuring mass?*" or "*There are 1000ml in a liter. How many ml are in one liter?*" right after Clark wrote down the vocabulary on the board) or reviewing the details from previous lessons (e.g., "*what did we use to balance the scale?*"). Although these questions were categorized as DF talk, students were only asked to recall factual information. On the other hand, in Sarah's or Margaret's class, such memorization questions were also present, but to a lesser degree. Sarah's and Margaret's DF questions offered various opportunities to engage students in different kinds of thinking, such as recalling the measurement unit on a liquid bottle or describing experiences related to ice.

Taking the factual questions about standard units as an example, all three teachers asked students to recall the standard unit of length/weight/volume. Clark first told students that "*our standard unit of measuring mass is gram.*" Then he reiterated the same concept with a simple question "*what's our standard for measuring mass?*" Such questions were simply repeating the information he presented earlier and the way he was asking did not include much related information about the standard unit to help students connect with other measurement concepts.

While Margaret also asked about the standard unit, her initiated question was "*what unit do I use to measure mass or how heavy something is?*" And then she continued her question with "*what tool did I use to measure mass or how heavy something is?*" Her questions reminded students conceptual understanding about measurement (i.e., corresponding properties of the measurement units) rather than simply addressing one measurement fact.

In Sarah's class, when she reviewed that a vial is not a good way to measure, she then asked "*what's the standard unit?*" and elaborated it with another question, "*when you buy liquids, what do we have on the bottle?*" She provided useful contexts for using the standard unit in her questions to help students recall the information and make connections to the context and the factual knowledge. Both of Sarah's and Margaret's questions were qualitatively different from Clark's recalling questions. Sarah and Margaret incorporated related information to their questions to help students connect to those associated concepts or contexts but such connection was absent in Clark's instruction.

In summary, from the descriptive analysis and examples of the three teachers' initiated questions, Clark's talk was mostly about procedural knowledge, and his DF talk was rather simple and lacked connecting concepts to help students recall or make sense of the factual information. Although Sarah's and Margaret's talk also contained a certain amount of procedural knowledge, their talk about DF knowledge not only asked students to recall factual information but also helped students make connection to the relating concepts or providing contexts for the use of such information. These variations on teacher enactment of the curriculum provide important information when one is studying students' learning of scientific inquiry or evaluating the implementation of science education reform.

4.2.1.2 Teacher talk during the sound module

Due to the lack of class observations of Clark's class during the sound module, only the results of Margaret's and Sarah's class were presented in Table 11. In the sound module, the most frequent talk was DF talk with an average of 61% in Sarah's class and 71% in Margaret's class. More CE talk was involved in both Margaret's and Sarah's classes than in the measurement module. Sarah's class had slightly higher percentage of CE talk (19%) than Margaret's class (10%). Overall speaking, the two teachers shared similar distribution on the DF, CE and Proc talk.

Table 11. The distribution of different types of teacher talk on the sound module.

	Sarah	Clark	Margaret
DF	61% (0.191)	n/a	71% (0)
CE	19% (0.206)	n/a	10% (0)
Proc	14% (0.125)	n/a	19% (0)
Other	6% (0.054)	n/a	0% (0)
Total units	20.67	n/a	21

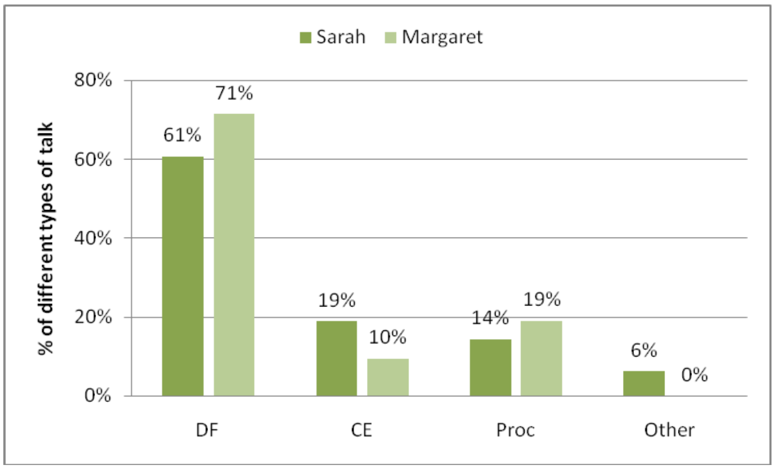


Figure 7. Comparison of different types of teacher talk on the sound module.

Unlike the measurement module where teachers did not provide many opportunities to explain (CE talk) after engaging students in the hands-on activities, in the sound module, both teachers engaged students in observing scientific phenomena, talking about their observations, or exploring the patterns of the phenomena, and also went on to explain the patterns of sound, predict the pitch of sound or extend their observations to other similar phenomena. For example, in Sarah's class, she invited a student to demonstrate bowing her violin in one class. Sarah helped students explain that by pressing on different positions of the violin string to make different sound, she was actually bowing the string of different lengths. By including a musical instrument as an example of sound phenomena, Sarah helped students connect the patterns of sound to musical instruments. In Margaret's class, she also stressed a concept that sound is caused by vibrations by asking students to feel the vibrations of their vocal cord. The occurrence of CE talk in both Sarah's and Margaret's classes indicated that both teachers provided opportunities for students to understand the related big ideas of sound.

Although both teachers aimed for conceptual understanding, in the observed classes, Sarah was more focused on students' daily life experiences by creating scenarios of sound phenomena. When Sarah first introduced the sound module, she intentionally dropped something on the floor while students did not pay attention, and then asked students how they knew she dropped things if they did not see it. Within this context, students were involved in discussion such as using ears to hear sound, sound can travel all over the places, etc.

In contrast, Margaret was more focused on the memorization of the explored patterns of sound phenomena and the clarification of students' confusion about sound. She helped students notice the same rhyme by asking that "when you have that *low* pitch sound, you have *low* vibrations, and when you have high pitch sound, how did they feel?" Margaret pointed out the similarity on pronunciation by mentioning that "when [the tongue depressor] is *lo*-ng, you hear a *low* sound." Additionally, Margaret was concerned about students' confusion about sound. In order to clarify the difference between pitch and volume, Margaret added a small activity for students to press their vocal cord while practicing making sounds of high pitch like mouse squeaking, and sounds of low pitch like bear grumble, with louder and quieter voices. In Margaret's class, students were provided with opportunities to memorize the patterns about sound.

The analysis of teachers' DF/CE/Proc talk did not provide a detailed picture of each teachers' enactment. Thus, the following section was to provide brief descriptions about the features of each teacher's class.

4.2.2 Descriptions of the talk from each teacher

In addition to the quantitative analysis of teachers' initiated questions, which may leave out a lot of information about what's going on in a classroom, more detailed descriptions about each teacher's instruction and class management were to provide a more realistic view of the learning environment of these three elementary science classes. The following descriptions started with two comparison teacher, Clark, whose school had similar SES background as the target teacher, and Margaret, whose school had higher SES background than Clark's school. The target teacher, Sarah, was presented last in order to compare with the descriptions of Clark's and Margaret's classes. For each teacher, three aspects of their science instruction were highlighted to briefly illustrate teachers' practice. Those included: (1) the flow of classroom activities, (2) the use of instructional tools, and (3) attributes of teacher talk.

4.2.2.1 Comparison teacher #1: Clark

Clark is a science specialist teacher who teaches grade K-6 in an urban public elementary school. The school enrolls 94% of African American students and 2% of White students, and 93% of the students qualify for free lunch. Clark has been a teacher for seven years and has been a science specialist teacher for four years in the same school.

(A) **Flow of classroom activities**. In a typical day of Clark's classroom, he usually began when he wrote down a few words as the objectives of today's lesson on the side of the blackboard and briefly summarized what students were going to do. During the class, Clark explicitly told his students to follow his instructions but rarely provided any visual support to help students remember those rules. Toward the end of the class, Clark would conduct a whole class discussion about today's activity and correct students about procedures that they did wrong.

(B) **Use of instructional tool.** Clark seemed to value student's motivations for doing hands-on tasks by adding extra objects for students to practice how to use a balance scale. Clark also brought an electronic balance scale to class to demonstrate a more accurate measure of mass. Clark's efforts to provide additional objects which were physically appealing to students did draw students' attention temporarily, but such attention was not directly related to scientific phenomena or was not followed by instructions that led to scientific thinking. Students' interest usually faded rapidly. When students were distracted or not following Clark's instructions, Clark had to stop the ongoing activity for class management, which made it even harder to focus on the big ideas of those activities.

Clark particularly emphasized the accurate order of measurement procedures during his instruction. For example, he prepared cards each printed with one step of procedure about using the balance scale. He called on different groups of students to sort out the cards in the correct order before giving out balance scales for students to work on measuring different weights.

Among the instructional tools that Clark added in the observed classes, none of them were directly related to the support of big ideas or critical concepts of the unit. Although the procedures of conducting the hands-on activities were faithfully implemented, the scientific thinking behind the activities was not supported in any form.

(C) **Teachers' talk.** The coding of Clark's talk showed that he usually initiated conversations to elicit simple factual knowledge such as vocabulary or simple questions which only required a yes and no answer. For example, when Clark reviewed hands-on activities, he only asked about the procedures for the activities (e.g., "*what did we use to balance the scale?*" "*Paper clips.*"), instead of the underlying concepts of the activities (e.g., "*What did we find out about using different paper clips?*"). In addition to questions for simple factual vocabulary, Clark also tended to ask yes or no questions. Clark asked questions like "*can I put paper clips on both sides of the scale?*" or "*can I put anything on the arm of the scale?*" or "*is it fair if I use paper clips to measure?*" Some questions may seem easy to answer with yes or no answer, but some questions required more than yes or no answer to respond and explain their thoughts. These simple-answer questions were not followed up with more elaboration on the related ideas and thus could be problematic on engaging students to appreciating scientific inquiries and guiding students to understand the connecting science content.

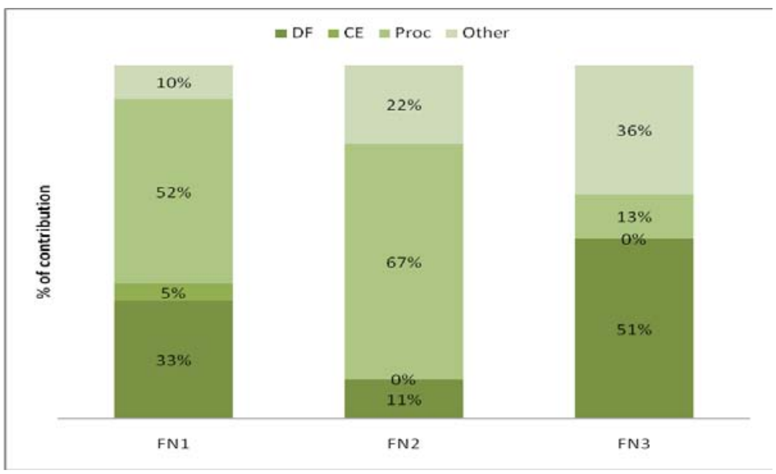


Figure 8. Proportions of the DF/CE/Proc talk of Clark's classes

4.2.2.2 Comparison teacher #2: Margaret

Margaret is also a science specialist teacher who is responsible for K-5 science. She has been teaching for 9 years and has taught for 4 years as K-5 science teacher. Margaret's school is an urban public school which composed of 37% of African American and 57% of White students. Looking at the teaching experience, Margaret and Clark shared similar teaching experience as science specialist teachers. But Margaret's school SES is somewhat higher than that of Clark's school.

(A) **<u>Flow of classroom activities.</u>** Margaret usually started her class with a review of the previous class, reviewed related vocabulary, or asked students to report on their science homework. During the class, Margaret would use the blackboard or transparency projector to remind students the procedures of the upcoming activities or help students organize the current concepts, and demonstrate how each procedure should be done. While students were conducting their investigations, Margaret would go around to different groups of students to monitor their progress or help with their questions. At the end of a class, Margaret usually asked different groups to write up their results on the blackboard and discuss the similarities or differences among groups. On some occasions, the class was ended in a hurry by having students clean up their materials and equipment without discussing the results with the whole class if Margaret lost track of time.

(B) **<u>Use of instructional tools.</u>** In Margaret's class, the instructional tools she used emphasized both procedures and ideas about certain activities. The tools that she used were usually aimed at organizing her instruction or students' concepts rather than at testing students' memory. For example, she used the transparency projector to present step-by-step procedures for the ongoing activity so that each group could refer to the written procedures during their hands-on activities. Margaret was concerned about whether students can come to a conclusion about what had happened at the end of the hands-on activities. She often reminded and asked students what are the current ideas and procedures of these activities using the blackboard or transparencies. For example, when Margaret started an activity to measure the volume of water, she briefly reviewed previous lessons about the standard units and tools for linear and weight measurement. She then found out that her students were not able to answer what units or tools were used. She drew a chart on the blackboard with columns of *length, mass and volume*, and rows of *what we measure, units we use and tools we use*. The chart helped students organize their ideas about measurement and connect ideas among different activities, instead of conducting activities and following measurement procedures step by step without making any connections.

(C) **<u>Teacher's talk.</u>** Margaret's initiated conversations emphasized both procedural and factual knowledge, depending on which class was observed. When students had finished their hands-on activities in previous lessons, the next class would involve more DF and CE knowledge; whereas if the class was about conducting an investigation, Proc knowledge would become a large proportion in Margaret's talk.

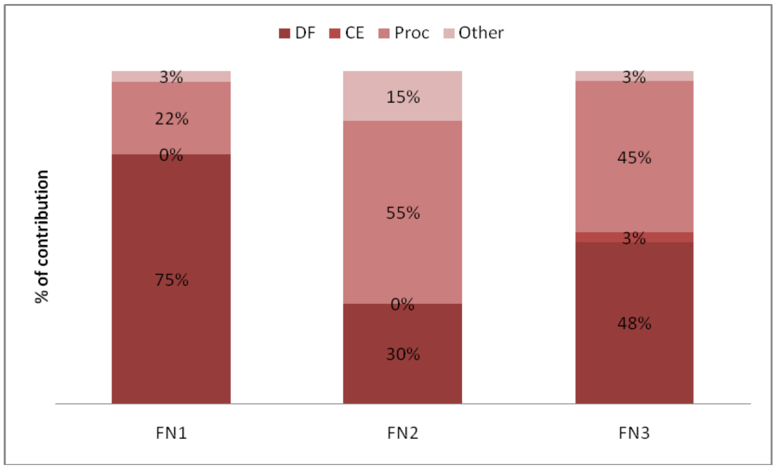


Figure 9. Proportions of the DF/CE/Proc talk of Margaret's classes

4.2.2.3 The target teacher: Sarah

Sarah taught third-grade science and math in a self-contained classroom. She had six years of teaching experience in science. Sarah's school is also an urban public school with 97% of African-American and 1% of White students. The school demographics of Sarah's was similar to Clark's school, but Sarah is not a science specialist teacher.

(A) **Flow of class activities.** From the observed classes, Sarah usually situated her students with a contextualized problem that they have to solve or she would create a scenario to direct students' attention to daily life phenomena. For example, at the beginning of a measurement class, she held a can of pop and asked students what they might measure about it (e.g., its length, weight and volume). Or she would point to the printed volume on the pop can and ask students if there was the exact amount of liquid inside the pop can so that they were not cheated. Or at the beginning of the sound unit, she made a loud noise by dropping something purposefully and asked her students how they knew that she dropped something if they were not looking. By embedding the upcoming scientific activities with daily life problems, students could make sense of the ongoing activity and apply their scientific knowledge.

During the class, Sarah usually demonstrated the procedures of the activities before students started their own investigation. Sarah would circle around different groups to help students with their emergent problems, monitor their progress or to demonstrate the procedures again. At the end of the class, Sarah would call on students to present what they observed, present their results, or explain their results. Sometimes Sarah would involve the entire class in discussions about the explanations of scientific phenomena and providing students with opportunities to evaluate different explanations.

(B) **Use of instructional tools.** The instructional tools that Sarah used also emphasized more about science concepts than procedural instructions. Those emphasized science concepts included the implicit ideas originated from the hands-on activities or the connecting ideas across similar activities. The tools which were used in the classroom included storybooks or problem-solving worksheets provided by FOSS, charts or tables created by Sarah with summaries of students' ideas or observational results, and real musical instruments which related to the ongoing sound concepts. My observation of how Sarah used these tools indicated that Sarah was focused on how students made sense of the measurement activities and how students explained different sound phenomena. Her students had to work on contextualized problems which require appropriate measurement procedures, or to generalize from the patterns of sound observations by summarizing the results of observations in a chart. The activities that students were engaged in were similar to scientific inquiry where scientists start with a problem they encountered or start with an interesting phenomenon they observed, design or plan systematic observations regarding the problem or research questions, then explore different patterns from their observations or organize their data in different ways to reveal patterns. Through the exploration of patterns, scientists could reason with their knowledge to find plausible explanations. The goals of scientific inquiry and the goals in Sarah's classes were more coherent in terms of rooting scientific activities in daily-life or actual problems to solve or to make sense of those observed phenomena. The instructional tools that Sarah incorporated in her class were mostly aimed to enhance students' understanding of science content or scientific inquiry, and less targeted to accurately following predetermined procedures.

(C) **Teachers' talk.** The coding results showed that every class provided a considerable amount of DF knowledge. The ways that Sarah approached students' talk about procedural knowledge were sometimes different from the other two comparison teachers. Many times when Sarah introduced students with hands-on activities, she engaged students in discussions about what procedures they could take to solve the problem at hand, and encouraged different groups of students to employ one method that they agreed on to conduct their investigations. Students not only had the opportunity to justify their proposed methods, but also to compare or evaluate their results with different methods from other groups.

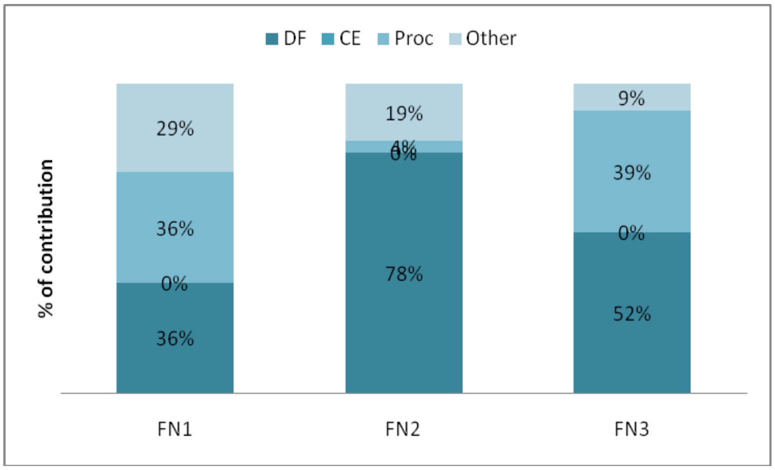


Figure 10. Proportion of the DF/CE/Proc talk of Sarah's classes

For example, one of the FOSS measurement activities required students to measure the volume of a soda can. Sarah introduced this activity to students by giving them a scenario that Ms. Schaeffer (a teacher from their school) is drinking slim fast and she is wondering if she gets her money worth since the can is written 325ml but the drink inside it is not filled up to the top. Sarah asked her students if they had any solution to the problem. After the discussion, students came up with two solutions: one is to fill up the slim fast with water and measure how much volume the water is; the other is to measure 325ml of water first and dump it in the can to see if it fills up the entire can. Sarah told her students to decide which solution they want to use as a group then they could start the activity. Sarah did not give out directions to students about what to measure but asked students to think about how they could use measurement procedures to resolve a daily life problem. By providing students a scenario or a purpose for the upcoming hands-on activity, Sarah intended to help her students make sense of the scientific inquiries and also the big ideas or the content knowledge related to the ongoing activities.

Such talk is somewhat different from Margaret's talk because Margaret rarely connected hands-on activities with real-life problems. Nevertheless, they were both concerned about students' content knowledge and helped students organize their thinking with tools such as charts or tables. Clark's talk seemed to differ from Sarah's and Margaret's talk in terms of the focus of the instruction, the talk he initiated, and the tools he acquired as aids to his instruction. Clark emphasized the procedures of conducting activities rather than the organizing concepts. He was concerned with engaging students' attention about conducting scientific investigations by including interesting manipulatives or phenomena but he did not lead discussions about how to explain scientific phenomena.

In summary, the percentages of teachers' DF, CE and Proc talk did not reveal much difference among the three teachers. However, in terms of the quality of teachers' talk, Clark's instruction looked different from Sarah's or Margaret's instruction. That is, Sarah and Margaret elicited questions that addressed connecting concepts between the ongoing activities and students' daily life experiences.

4.3 PATTERNS BETWEEN STUDENT LEARNING AND TEACHER TALK

Due to the lack of complete data on teachers' talk on the sound module, the analysis of the patterns between students' talk and teachers' talk were solely based on the measurement module so that the results were comparable across the three teachers. There was some incoherence between student talk and teacher talk because student talk in measurement module was categorized into CE and Proc knowledge, and teacher talk was categorized into DF, CE and Proc knowledge. The analysis of patterns between teacher talk and student talk combined DF and CE as one type of knowledge and contrasted with Proc as the other type to make comparison between conceptual knowledge and procedural knowledge. The results were summarized in Table 12. The itemized learning outcomes of conceptual and procedural knowledge were listed in Appendix J.

Table 12. Comparison between student talk and teacher talk among three classes.

Student talk* (% correct) Teacher talk† (% of contribution)	Sarah's class	Clark's class	Margaret's class
Conceptual talk (DF and CE knowledge)	**47%** 55%	**27%** 34%	**63%** 56%
Procedural talk (Proc knowledge)	**75%** 26%	**76%** 44%	**74%** 41%

* Student talk was based on the average percentage of correct responses across all items related to the same type of knowledge. Therefore, the total of student talk did not add up to 100%.

† Teacher talk was based on the average percentage of frequencies on teacher-initiated questions. The sum of teacher talk should be 100%, including conceptual, procedural and other talk.

4.3.1 Patterns of procedural talk

The results showed that for procedural talk (Figure 10), the performance of students' learning of Proc knowledge, as described in previous student learning section, was pretty similar across the three classes. Students' correct responses of the three classes ranged from 74% to 76%. But Proc knowledge of teacher talk was somewhat different, with the highest percentage occurred in Clark's (44%) and Margaret's (41%) class and the lowest percentage occurred in Sarah's class (26%). The results of procedural talk may suggest that exceeding a certain amount of instruction on procedural knowledge may not improve much on students' understanding. Students might not need as much procedural talk as provided by Clark and Margaret in the measurement module for students to master the measurement procedures. More time could be used to explore relating measurement concepts to ensure students' understanding on measurement.

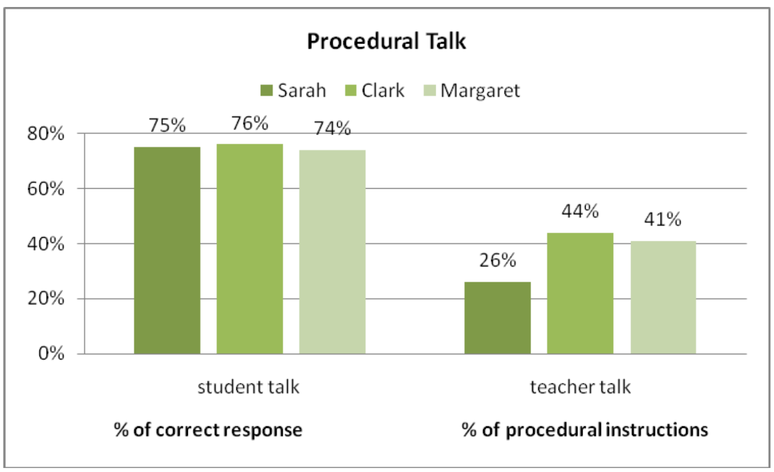


Figure 11. Comparison of procedural talk among the three classes.

4.3.2 Patterns of conceptual talk

For conceptual talk (Figure 11), students performed differently across the three classes and such differences also shown on the percentages of teacher talk. Sarah's and Margaret's instruction contained similar proportions of conceptual talk (55% and 56% respectively) among the three teachers, and their students also showed better performance on their conceptual knowledge with 47% (Sarah's class) and 63% (Margaret's class) of correct responses. About 34% of Clark's initiated-questions were related to conceptual knowledge, the least among the three classes, and the performance of his students also showed less understanding on the conceptual knowledge of measurement. The percentage of conceptual talk in teachers' instruction seemed to be positively related to correct responses on students' understanding of conceptual knowledge.

However, one thing should be noted that although Sarah and Margaret had similar percentages on their conceptual talk, Margaret's students performed better than Sarah's students, It showed that student understanding might also be affected by other factors, such as SES of schools, the types of concepts being emphasized, etc.

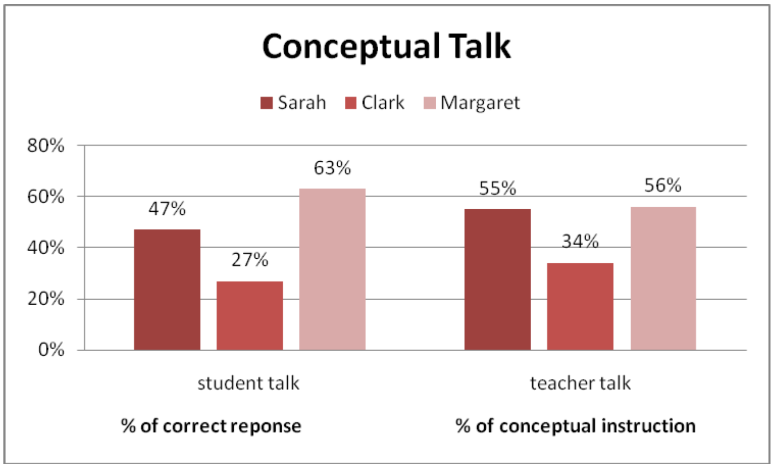


Figure 12. Comparison of conceptual talk among three classes.

4.4 COMPARISON BETWEEN DIFFERENT ASSESSMENTS

Did this study find differences among the various formats for assessing students' understanding? Two different formats of assessments (i.e., written assessments and performance interviews) were used in this study to evaluate students' learning outcomes. Although the design of the items was not intended to compare the coherence between the assessments, some of the tested concepts did overlap between the two assessments. To look at the interchangeability between the assessments, one needs a more systematic way to look at both assessments, such as parallel design of the items or use of multiple items around the same concept in both formats. Although the results of this study can not be used to guide the interchangeability of the assessments, we can compare some of the findings to begin to address this issue. In order to do that, the items which tested similar concepts with different forms of assessments were selected. Students from all three classes were aggregated to run the analysis.

The analysis of different types of assessments (Table 13.) was based on the percentages of agreement (for multiple-choice items) or correlation coefficient (for short-answer items). The percentages of agreement for each pair of items indicated the percentage of students who answered correctly in both forms of assessments plus those who answered incorrectly in both forms. Since the items were not parallel, item difficulty, which is the percentage of correct response, may vary a lot from item to item. Thus, some of the disagreements resulted from different levels of difficulty even for items desgined to test the same concept. For the selected items from the measurement module, the written assessments seemed to have higher percentages of correct responses on the items related to measurement procedures. The results showed that the percentage of matching conclusion from both assessments was about 45% to 73%. The correlations between short-answer items and explanations during performance assessments were all below moderately correlated (r<.40). Thus, given the low correlations, these items may be testing different skills. Students who were more articulate or wrote better may have advantage on a certain type of assessment.

Table 13. Comparison of the difficulty level and item correlation on selected items from the written and performance assessments.

	Items on written assessment (% correct or mean(std))	Items on performance assessment (% correct)	Comparison (% consistent or correlation)
Measurement procedures:			
Choose properties of interest	86%	49%	46%
Choose ruler to measure length	86%	69%	69%
Measurement concepts:			
Estimate length appropriately	70%	81%	73%
Choose same-sized units to measure	58%	55%	45%
Properties of sound:			
Identify tautness_1	38%	42%	50%
Identify tautness_2	35%	42%	61%
Processes of sound:			
Sound production	Mean = 1.1 (std=0.47)	1.07 (0.83)	r=.189
Sound transmission: thru air	1.10 (0.83)	2.20 (0.90)	r=.322
Sound transmission: thru string	0.35 (.049)	3.26 (1.5)	r=-.105

Two of the measurement items were parallel in terms of tested concepts and difficulty level, but the results of these two items also showed moderate percentage of consistency. Those two items[4] both targeted on the concept of uniformity and asked students to choose an appropriate tool to measure a certain length among three choices of tools. In the written assessment, the item was presented with a drawing of a jump rope as the object to measure, and three choices of tools were drawn below a jump rope for students to choose from. The correct answer was pencils which were all the same size, and the distracters were different-sized blocks and broken crayons. In the performance assessment, students were asked to measure the length of a book with either same-sized batteries, broken crayons, or different-sized screws. These two items were parallel in terms of the structure of the problem and the distracters of the problem. Their item difficulties were 0.58 and 0.55, respectively. About 45% of students answered correctly on both items, which left 55% of students answering correctly on one item but not the other. From the comparison of these two items, it suggested that students' performance differed in the different assessments formats or that different assessment formats may be testing different abilities.

4.5 SUMMARY OF RESULTS

The results of the study were summarized in the following order: (1) comparison between the three classes on students' learning; (2) comparison between the three teachers on the enactment of FOSS; and (3) the relations between students' learning and teachers' enactment.

4.5.1 Comparison between the three classes on students' learning

In the assessments of measurement module, students performed similarly on employing appropriate measurement procedures for standard measurement tools, such as identifying the property of interest, placing tools appropriately, reporting accurate results, etc. But students differed on some of the learning outcomes which required conceptual understandings and explanations about measurement. The differences included making appropriate estimations with

[4] One of the items was from BI_3 of written assessment, and the other item was from task 1e of performance assessment. In the assessments, both choice of tools and rationale were asked. But the analysis presented here was only based on students' choice of tools in order to make the comparison more similar between the two items.

standard units and using identical units to measure things. Less of Clark's students provided reasonable length estimations than Sarah's and Margaret's students, which indicated that students had not formed a general mental representation of standard units of length.

Generally speaking, Clark's students were able to employ appropriate measurement procedures but did not show appropriate understanding of measurement concepts. Sarah's and Margaret's students showed better performance on their conceptual understanding of measurement than Clark's students. The results implied that although procedures can be taught by step-by-step instructions, the related conceptual understandings underlying these hands-on activities did not come naturally by simply replicating the exact measurement procedures.

In the sound module, students showed differences in their understanding of sound properties and sound processes. Most of the differences occurred between Clark's and Margaret's students, including identifying the factors affecting sound pitch, and explaining the processes of sound production and sound transmission. The performance of Sarah's students was mostly in between that of the students of the two comparison teachers. More specifically, Sarah's students performed better than Clark's students but less well than Margaret's students, though not significantly different from either of the classes. The assessed learning outcomes were related to students' conceptual knowledge and thus confirmed the results from the measurement module that students varied in their conceptual understandings.

4.5.2 Comparison between the three teachers on the enactment of FOSS

The three teachers provided different instructional tools to support procedural and conceptual learning in the measurement module. Clark focused on having students follow the instructed measurement procedures, whereas Sarah and Margaret focused on connecting the ongoing activities with related concepts or daily life experiences. More specifically, Clark provided instructional tools to help students memorize the order of the measurement procedures by having students sort out the order. On the other hand, Margaret and Sarah constantly made connections between activities' procedures and underlying concepts. Margaret helped organize related measurement concepts by constructing a chart to remind students about the relationship between measurement tools, units, and properties of objects. Sarah provided meaningful purposes to engage students in scientific activities by creating daily-life problems for students to solve.

The types of questions that teacher initiated in the classroom indicated that most inquiry opportunities teacher created through teacher-asking were regarding descriptive or factual knowledge (e.g., describing observations, eliciting factual results, reviewing previous activity procedures, or recalling or explaining vocabulary). In contrast, they rarely asked students to rationalize the ongoing procedures, to explore the general relationships behind scientific observations, or to evaluate different solutions or explanations. The occurrences of CE talk were rare across three teachers. The distribution of teachers' talk is probably better estimated by recording the amount of time teacher spent on each type of talk. However, the questions initiated by teachers still provided information on the type of thinking that teacher focused on and interacted students with.

In spite of the similar numbers shown on the proportions of teachers' DF, CE, Proc talk, the format of teachers' questions revealed some differences among the three teachers. Clark's questions mostly prompted students for simple and immediate answers, such as recalling factual knowledge, or repeating what he just said. Margaret's and Sarah's questions required more explanations and usually served purposes such as scaffolding or asking students to elicit past experiences, to reflect on the difficulty or easiness of the hands-on activities, to provide rationale for their solutions, etc. These differences in teachers' enactment affected the types of inquiry opportunities provided for students and different focus of learning.

4.5.3 Relations between students' learning and teachers' enactment

Combined the DF and CE knowledge as conceptual talk in teachers' instruction, the patterns between student learning and teacher talk showed that when teachers provided more support for students' conceptual knowledge, students also showed better performance on conceptual knowledge. Clark's questions indicated less support for students' conceptual understanding than Margaret's and Sarah's instructions, and the results also reflect such relation on students' learning outcomes. On the other hand, although students' procedural performance was not different among the three classes, Sarah's procedural talk was less than Clark's and Margaret's instructions. It indicated that when teachers provided less support on conceptual understanding and more support on procedural practice, students may perform equally well on their procedural knowledge but show less understanding of conceptual knowledge.

4.5.4 Comparison between different types of assessments

Some of the items were selected to compare the results between written assessments and performance assessments. Among the items that were selected to test the same concepts, only moderate consistency or correlation were found between two assessments across all selected the items. Two of the items from the measurement module that tested the same concept and had similar difficulty level also produced moderate consistency between written assessment and performance assessment. Different types of assessments can be testing different types of abilities.

4.5.5 Problems of type I error

Since this study compared different assessment items on two modules, it could bring up concerns for type I error, which means the rejection of null hypothesis when there is actually no difference between two samples. In this study, thirty-five statistical tests were conducted throughout the comparison of the three classes. Among which, eleven results showed significant differences. Since the alpha value chosen for this study is .05, the chance of type I error is .05, which indicates that 5% of the tests may be significant simply by chance. For example, running 100 statistical tests may yield 5% of tests to come out significantly different, but actually one sample did not differ from the other. There were a total of 35 statistical tests run in the analysis of this study, so the possibility of type I error was 1.75. There were eleven tests showed significant differences and minus the possible 1.75 tests which might be due to chance, there were at least nine tests that indicated the differences among the three classes. Therefore, given the possible type I error, the significant differences were still more than the possibility due to random error.

5.0 DISCUSSION AND CONCLUSION

5.1 INTRODUCTION

As noted in chapter one, the adoption of hands-on science curricula has brought new challenges to teachers and students (Marx et al., 2004; Schneider & Krajcik, 2002). Recent studies have focused on the impact of using hands-on science curriculum with regards to teacher education or student outcomes separately, but few studies consider both teachers and students. This study used multiple assessments to evaluate student learning and explored the use of a hands-on science curriculum in elementary science classes through the examination of students' learning and teachers' enactment of the curriculum. One target teacher participated in a professional development (PD) workshop, PPHOSS, which targeted the support of teachers' inquiry-based instruction using a BITT instructional model (Cartier, 2003). Two comparison teachers who had similar teaching experience but did not participate in such workshops were also included. The three participating teachers focused on different aspects of the scientific activities while enacting the curriculum. As for students' learning, the results were varied on their learning of conceptual and procedural knowledge.

More specifically, the results of student learning showed similar performance in the use of measurement tools, such as aligning objects with the origin point, balancing a scale, or choosing a graduated cylinder to measure liquids. However, the differences among the three classes appeared in their conceptual understanding of measurement and sound, such as estimating measurement with standard units, or explaining how sound travels. The results indicated that although students were able to demonstrate appropriate measurement procedures, they may not have fully comprehended the rationale behind using measurement tools. The manipulation of measurement tools can be easily mastered without appreciating the importance of using standardized units or estimating measurement results. The explanations of measurement procedures or scientific phenomena (e.g., sound production or transmission) do not come automatically by following procedures of hands-on activities or reproducing phenomena. Student learning of conceptual knowledge requires much more guidance from the teacher to elicit students' ideas and direct students' attention exploring patterns, constructing plausible explanations, differentiating supporting and non-supporting evidence, or communicating thoughts with others. The difference in students' understanding of science concepts among the three classes should bring attention to educators and researchers to further support the probing of students' ideas and explanations being the central elements of hands-on science activities.

The findings of teachers' enactment of the curriculum revealed differences in the content of the questions being asked in the three classes. Sarah's and Margaret's questions showed higher proportions of descriptive/factual (DF) talk (or conceptual talk) than Clark's in the measurement unit. Clark initiated more procedural (Proc) talk instead. Taking a closer look at the format of teacher questions, Clark's questions were attempts to elicit simple agreements from students such as yes or no, whereas Sarah and Margaret often connected their questions to the related concepts of the ongoing activities or attempted to elicit explanations of the phenomena or procedures. The BITT model suggested in the PPHOSS workshop emphasized the planning of lessons around the central science concepts of a focused module and making connections across different activities and different modules. Sarah's instructional questions provided evidence of her emphasis on conceptual understanding, which may have resulted from her mastery of the BITT model.

Moreover, the format of teacher questions was also different among the three teachers. Teacher questions provide different types of inquiry opportunities to students and thus may enhance or limit the thinking processes that students engage in. Sarah's and Margaret's questions were often open-ended questions, such as explaining the rationale of their solutions to a problem, describing their experiences with ice, or summarizing science stories from reading materials. Those questions often required more explanations, memorization and reflection, rather than pure guessing. On the other hand, Clark's questions were rather short and simple without giving students much time to think or reflect. Such differences in teacher questioning may contribute to the differences in students' conceptual understanding.

Linking the results of student learning and teachers' enactment of the measurement module, there was a positive relationship between students' conceptual knowledge and teachers' conceptual talk. That is, higher percentages of teachers' conceptual questions were associated with more accurate responses by students on the items regarding conceptual knowledge. For procedural talk, since there was not much variation in students' performance, the pattern between students' learning and teachers' enactment was not shown. It appears that the focus of instruction (i.e., science concepts or procedures) directs students' attention to teachers' instructional objectives and may have a direct impact on what students learn. The FOSS curriculum may have already provided sufficient direction on conducting investigative procedures, but more instructional information about science concepts may be necessary to help students conceptualize the underlying science content.

5.2 LIMITATIONS AND SUGGESTIONS

Although the three participating teachers displayed some differences in their enactment of the curriculum, it should be noted that several factors outside the scope of this study may also contribute to their differences. For example, the teachers were selected either by the district supervisors or from participating in a PD workshop. The small sample size may not be representative of the larger population of public school teachers in the United States. The selection of teachers does not necessarily represent their schools or other schools in similar socioeconomic status (SES) communities. Additionally, the teaching portfolios of the three participating teachers were not exactly matched. The comparison teachers were science specialists whereas the target teacher taught in a self-contained classroom. Science specialists are usually responsible for science classes from kindergarten to grade five and are sent out for science-relevant professional development by schools. The differences in their teaching responsibilities may affect their view of effective teaching. Science specialists may acquire more advanced pedagogy because of their rich experiences of attending science workshops. Teachers of self-contained classrooms may have more opportunities to make connections on scientific phenomena in students' daily lives because those teachers share an extensive amount of time with students rather than a few hours per week.

Differences such as school SES and teachers' teaching responsibilities were not controlled in this study due to the small sample size. However, the aim of this study was to demonstrate the possibility that teachers may enact a given curriculum differently, and this might affect their students' learning. The correlation found in this study between high conceptual teaching and high conceptual learning may not be replicated in other studies. More samples of teachers and students and more thorough observations will be required to further evaluate the enactment of hands-on curricula in actual classrooms and its relation to student learning.

In addition, this study employed a post-test only design. The differences among the three teachers could reflect the impact of PPHOSS, or it could also result from other factors not examined in this study. Since the scope of this study was set to include factors within the classroom level, there remain other factors beyond the classroom level that may still play a significant role in student learning and teacher instruction. For example, Sarah's school seemed to be very different from Clark's school. These two schools share similar SES distribution among students, but showed differences in teachers' instruction and student learning. The unwritten rules or decisions underlying a school, such as what types of behavior are stressed or how students or teachers were treated, inevitably and unnoticeably shape the interaction between students and teachers. Such factors were beyond the focus of this study, but should not be ignored while interpreting the results.

Moreover, the results of student learning outcomes and teachers' enactment may be affected by the selected science modules. The focused modules of this study were measurement and physics of sound. The measurement module does not reflect the typical cycle of scientific inquiry as much as the sound module, where students start from exploring phenomena, observing patterns, finding explanations, and evaluating their explanations, as illustrated in Figure 1 of chapter two. In the measurement module, the lesson goals indicated in the teacher manual are to practice measurement procedures and to memorize standard units without much reference to their connection with scientific inquiry or explanations. If teachers simply followed the written curriculum without considering the need to connect with big ideas in science, then teachers' enactment may turn out to be very procedural-oriented. Thus, the instruction of the measurement module may highlight teachers' focus on procedures more than other modules. Although other modules may not signify the procedural vs. conceptual enactment as much as the measurement module, what types of knowledge teachers tend to focus on will still be reflected in other modules, albeit perhaps to a lesser degree.

Furthermore, the search for a coherent coding scheme that links student learning and teachers' enactment may limit its sensitivity to other characteristics of teachers' instruction. This study used a coherent coding scheme focused on the types of science content knowledge for both student learning and teachers' enactment in order to explore the connections between these two aspects. Thus, the coding scheme might miss the differences in the depth of inquiry opportunities provided in the instruction.

For example, the three participating teachers in this study did not differ much on their percentages of DF, CE or Proc talk in class, but the way they asked students questions were very different. The questions in Sarah's and Margaret's classes were more focused and clear on the thread of concepts that the teachers want to address. They also created more questions by following up students' answers or elaborating their original questions. On the contrary, Clark's questions were either rather unclear on the intended concepts or very simple which mostly required yes/no or dual-choice answer. Clark seldom followed up on students' ideas or elaborated his questions when students did not respond.

Such differences in teachers' enactment were not captured if only science content knowledge were coded. Several possible codes to include in teachers' enactment can be teachers' follow-up questions around the same topic, the expected answers of teachers' questions, the occurrence of no-response questions, etc. The inclusion of such codes may provide more information on how science concepts were guided by teachers or accessed by students. Future studies could pay attention to not only the covered content of teacher talk, but also the way that such questions were addressed. Other modeling techniques of statistical methods (e.g., path analysis, or structural equation modeling) may be needed to examine the correlations among these variables holistically if larger samples are used. Additionally, videotapes and transcriptions of classroom talk would be needed to do these analyses and further examine the function and depth of teacher talk. Field notes taken in the classes only contain limited information in a limited amount of time. For a more complete analysis of teacher-student talk in the classroom which targeted on teachers' reactions to students' responses, videotaping of the classes would be required.

The comparison between different formats of assessment resulted in moderate percentages of agreement and correlations. This finding suggests that different assessment formats could affect student performance outcomes. Therefore, multiple assessments may be necessary to adequately evaluate students' understanding of scientific inquiry and science knowledge. Although performance assessments seem to be more consistent with the desired outcomes detailed in the National Science Standards, there remain challenges of administering, scoring and interpreting those assessments. More studies of the development of performance assessments, as well as their reliability and validity, are necessary to build clear connections between the results of assessments and students' understanding of scientific inquiry.

5.3 IMPLICATIONS FOR PROFESSIONAL DEVELOPMENT

The variations in teachers' enactment showed us how the same curriculum can be interpreted differently by teachers who are not provided with specific support targeted to engaging students in scientific inquiry while enacting hands-on science curricula. One might conclude that not all teachers emphasize the essence of scientific inquiry, and that some teachers might consider the sole focus on the procedures of the activities as implementing science curriculum faithfully. Without a clear focus on the central science concepts, teachers may want to maintain the order of the class by encouraging students to follow the exact steps of instructions to conduct scientific activities, rather than initiating discussions or guiding students around science concepts related to those ongoing activities. If teachers do not understand the rationale for engaging students in those hands-on activities, it is unlikely that their students will learn much science by following the procedures for conducting activities. Therefore, it is urgent to provide teachers critical support on the enactment of the curriculum.

5.3.1 The emphasis on teachers' conceptual talk

From the results of this study, there exist discrepancies in students' conceptual understanding among the three classes. How should educators better support the development of science concepts in the classroom when using hands-on science curricula? Researchers who took the stance of sociocultural theories of learning emphasized the importance of supporting teachers in different ways, such as through professional development (Putnam & Borko, 2000; Schneider, Krajcik & Blumenfeld, 2005), or discourse communities for teachers (Ball, 1994; McLaughlin & Talbert, 1993; Putnam & Borko, 2000). Several features of the PPHOSS workshop may have helped in supporting teachers' conceptual talk.

In PPHOSS, Cartier (2003) highlighted the use of an instructional model, BITT, to assist teachers on structuring their lessons. The BITT model supported and emphasized the development of conceptual understanding by bringing in three key elements of inquiry-based instruction—Big Ideas, Tools and Talk. These elements were to help students make sense of scientific activities and construct science concepts. The identification of Big Ideas within each module could help teacher plan the instruction around and make connections among their targeted science ideas instead of being distracted from the ongoing hands-on procedures. The inclusion of instructional Tools created opportunities to organize, structure, or recall related experience and science concepts. Engaging students in Talk may facilitate students' skills in communicating their ideas and informs teacher about students' learning.

Sarah's enacted curriculum seemed to incorporate the BITT model nicely by starting from a few big ideas that she included in her lesson plans (e.g., "things have different properties," "tools to measure different properties"). She used concept maps and charts to connect different activities within the same module around the big idea of "things have different properties." She also asked questions to invite students to explore patterns of pitch and length, to come up with a solution to decide whether there were the same amount of lemonade in different containers, to rationalize their solutions for examining the capacity of a soda can, and to explain how a violin makes different sound pitches. Sarah created and employed inquiry opportunities during her instruction to prepare her students with authentic scientific practices. In this example, the insertion of the BITT model into hands-on science activities enhanced Sarah's instruction of conceptual knowledge and thus may be helpful in promoting students' learning of science concepts.

Moreover, other features which were introduced in Putnam and Borko's (2000) paper also supported the notion of situated learning and could be added in future PD studies to further enhance teacher learning. First, have researchers work with teachers in their own classrooms so that learning occurred in the classroom and teachers were situated in the context that is the same as their teaching environment. Second, use artifacts from the classroom during PD workshops. Artifacts such as student works or videoclips from real classrooms can be brought to group discussion so that the communities could share ideas, thoughts, techniques, strategies and suggestions with each other.

5.3.2 Implications for research in teacher education

From the limited samples included in this study, the PPHOSS workshop seemed to have produced a positive impact on Sarah's instruction. However, it is still unclear how the BITT instructional model influenced Sarah in her implementation of the curriculum, and the evidence of including elements of the BITT model in her instructions. Building connections between the BITT instructional model and teachers' enactment could help us better understand how BITT supported instructions about scientific inquiry in the classroom. Future studies could investigate the relation between teaching which incorporates the BITT model (with the training of PPHOSS or without the workshop) and students' learning. Other participants who had attended this PD workshop can also be included to investigate the use of the BITT model more thoroughly.

The impact of the PD workshop in the following years of Sarah's teaching could also be studied to find out the sustainability and appreciation of the BITT model. Teachers have to adapt to the changes in the national and state standards, the curriculum, and the policy, value and culture of their schools. It is impossible to implement a method which can easily fit in every context. Teachers always have to make their decisions and judgment to apply the changing rules to their own classrooms or to different cohort of students. Whether PD can be long lasting is one of the critical concerns of researchers. Future studies could follow up on the change and adaptation of the target teacher, or other teachers in the PD workshop, on their teaching in the following years and also in different modules to examine the impact of the BITT model.

Margaret, who did not participate in the workshop, also included a significant amount of conceptual talk in her instruction. Although additional information about Margaret's pedagogical beliefs or history of teacher professional training was not included in this study, Margaret was involved in another professional development project conducted by Jennifer Cartier. According to the data collected from that project and some unsystematic class observations in this study, Margaret outperformed her peers in teaching science in her school district. In an interview in which she talked about her view of effective teaching in science, she mentioned that before instructing a science unit, she carefully reviewed the entire instructional materials, including teacher manuals, students' reading texts, video materials, etc. From observing her class, she evaluated the feasibility of implementing the hands-on activities recommended in the curriculum and actively adjust the procedures or materials of the activities. This level of agency over the instructional materials is not common among elementary teachers. However, such agency may be necessary for teachers to construct lessons on the conceptual level rather than procedural level (Remillard, 1999; 2000; 2005).

Moreover, being a science specialist, Margaret might have richer opportunities to attend professional development workshops particularly related to elementary science instruction or to gain more access to professional communities about science teaching. At a school level, teachers are assigned to attend different workshops based on their teaching specialty and thus specialized teachers usually participate in workshops directly related to the content which they are teaching. Such experiences may also contribute to her effectiveness and thoroughness of teaching science.

5.3.3 Implications for research on student learning

This study also added to the literature on students' concepts of measurement and sound to inform researchers regarding the design of related hands-on activities and building on students' prior knowledge. The analysis of students' understandings of measurement showed that students may employ several purposes of measurement at the same time. In the tasks when students were asked to use nonstandard tools to measure, students were looking for different purposes of measurement, such as expedience (using longer tools), or flush-fit at both ends (using tools so that they with different sizes to cover both ends of the measured object). A more sophisticated purpose of measurement is to use identical units to produce reliable measurement across time and space. Younger children were concerned about less important purposes for measurement. About 32% of students were not consistent in showing the uniformity concept across different assessments. They chose tools to fit flush with both ends of the measured object or chose longer tools to expedite the measurement procedures as more important than tools with identical units. These different purposes of measurement may each reflect a practical yet less important aspect of measurement, such as accuracy and expedience compared with reliability. Students may have had to explore these competing purposes and gradually prioritize the important ones throughout the conceptual development of measurement (Lehrer, 2003).

Students' explanations of sound phenomena confirmed previous findings that children tend to view sound as a substance which contains several properties, such as being pushable, frictional, containable, consumable, locational, transitional, stable, additive, inertial, and gravity sensitive (Eshach & Schwartz, 2006; Reiner, Slotta, Chi, & Resnick, 2000). In this study, children also showed confusion about viewing sound as a substance instead of a form of energy. It was very common that children explained how sound transmitted through solids by describing that sound came out from the cracks between obstacles, whereas a more accurate description should be the vibrations of sound makes the solid object or air molecules vibrate so that the sound traveled. Some children also conceptualized sound having similar properties with air (also a substance) and said that "it spreads out everywhere, like wind."

Other forms of energy will be introduced in the magnetism and electricity module (4th grade), solar energy (6th grade) and the matter and energy module (middle school), after the sound module in the third grade. It seems reasonable for students to start noticing that sound has other properties which are not the same as substance. For example, energy can be transmitted through media but not matter. Students may explore the phenomena that sound can travel, and sound will not be restricted or blocked by solids, liquids, or gases. Researchers could design activities around discovering the differences between substance and energy. Other pre-conceptions such as sound travels like air, or sound travels because it is loud, can be considered as potential learning objectives for students to distinguish between sound and air and discover the difference between the two.

APPENDIX A

HANDS-ON SCIENCE ACTIVITIES CONDUCTED IN MEASUREMENT MODULE

Hands-on activities	What students do…	Science concepts
The First Straw	Students learn the need for standard units of linear measurement. They measure objects with nonstandard units, straws, and then use a meter tape to measure objects in meters and centimeters. Students measure and compare body dimensions in the metric system.	▪ standard units made comparison possible ▪ standard units of length include cm, m, km
Weight Watching	Students learn the need for standard units for measuring mass and use the FOSS balance and mass pieces to weigh objects. Students prepare 100-g bags of gravel and cooperate to make a kilogram mass piece. They discover that a sponge can soak up many times its own mass in water.	▪ standard units made comparison possible ▪ standard units of weight include g, kg
Take Me To Your Liter	Students learn the need for standard units of volume. They use syringes and graduated cylinders calibrated in milliliters to measure fluids accurately. After learning how to use the FOSS volume measuring tools, students measure the capacity (maximum volume) of several common containers.	▪ standard units made comparison possible ▪ standard units of volume include l, ml
The Third Degree	Students compare the temperatures of three cups of water using their fingers, which leads to the need for a measuring tool and standard units. Students use alcohol thermometers and measure in degrees Celsius. They measure the temperatures of warm and cold water and find out how cold a mixture of ice and water gets in 10 minutes. The module ends with a Metric Field Day as students compete and officiate in events designed by the class.	▪ standard units made comparison possible ▪ standard units of temperature Celsius

APPENDIX B

HANDS-ON SCIENCE ACTIVITIES CONDUCTED IN PHYSICS OF SOUND MODULE

Hands-on activities	What students do…	Science concepts
Dropping In	Students explore their ability to discriminate between sounds, by dropping objects into a drop chamber and identifying each object by the property of its sound. They develop a code by assigning letters to objects and send messages to one another by using their drop code.	▪ Objects make different sounds ▪ Sound can be used to transmit information
Good Vibrations	Students explore sound generators (tongue depressor) and musical instruments (xylophone tubes, calimba, door fiddle) in miniactivities to find out what causes sound and what changes the pitch. They investigate variables that affect changes in pitch: the length of vibrating objects and the tension on vibrating strings.	▪ Sound is caused by vibrations ▪ Longer objects make lower-pitched sound
How Sound Travels	Students work in collaborative groups on miniactivities that introduce a sound source and a medium of sound travel. They observe and compare how sound travels through solids, water, and air by listening through a wood stick, a desk, a string, etc.	▪ Sound travels through different mediums ▪ Sound is the loudest through solids than air
Sound Challenges	Students investigate the nature of our sound receivers, ears. They are challenged to put their knowledge of sound sources, sound travel, and sound receivers to work. They take one of the instruments they used earlier and change its pitch, make its sound travel farther, or make it louder.	(apply the sound concepts students learned earlier)

APPENDIX C

INSTRUMENTS USED FOR MEASUREMENT MODULE

Selected FOSS assessment items:

3. A good estimate of the length of this piece of paper is

   A. 3000 cm.
   B. 300 cm.
   C. 30 cm.
   D. 3 cm.

4. What is the best way to tell the exact temperature of a cup of water?

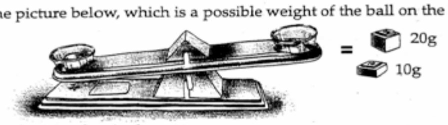
   A. Stick your finger in the water.
   B. Take a sip of the water.
   C. Put a thermometer in the water.
   D. Pour the water into a graduated cylinder.

11. A good estimate of the mass of this sheet of paper is

   A. 5000 g.
   B. 500 g.
   C. 50 g.
   D. 5 g.

14. In using the picture below, which is a possible weight of the ball on the left?

   A. 10 g
   B. 20 g
   C. 30 g
   D. 40 g

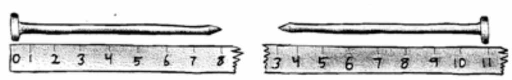


15. The height of my teacher is closest to

   A. 2 mm
   B. 2 cm
   C. 2 m
   D. 2 km

16. Look at these drawings. On the line next to the drawing, write out the measurement shown.

26. water __________



28. Use the picture above to answer the questions.

a. Compare the nails. Is one nail longer than the other? __________

b. How long is each nail? How did you get your answers?

____

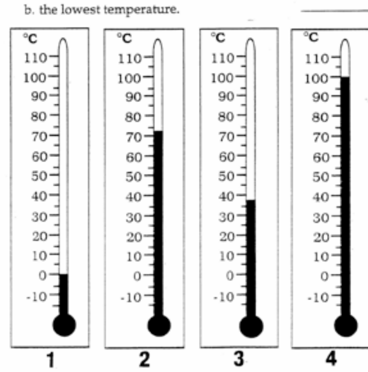
____

____

24. The thermometers numbered 1–4 on this page show different temperatures. Write the number of the thermometer that shows

a. the highest temperature. __________

b. the lowest temperature. __________

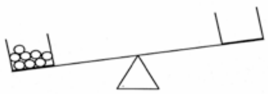


1 2 3 4

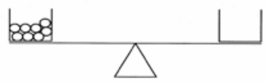
Big Idea Written Assessment:

1. I have a scale and two cups. I put all of my marbles in one cup, and nothing in the other one. What will my scale look like?

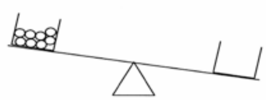
a.



b.



c.



Use some or all of the marbles to make this scale balanced. <u>Draw a picture</u> to show what you did.

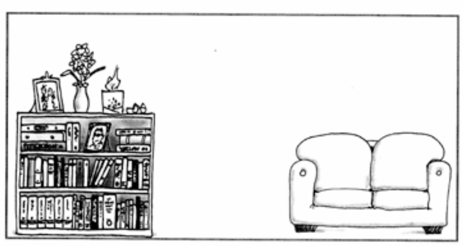
Why do you think your solution will work?

____

____

____

2. Sharice wants to buy a desk to fit between a book shelf and a sofa in her den. What should Sharice know about the desk to decide if it will fit?



a. how strong the desk is.
b. how tall the desk is.
c. how wide the desk is.
d. how heavy the desk is.

Please explain to Sharice why knowing this information would help.

____

____

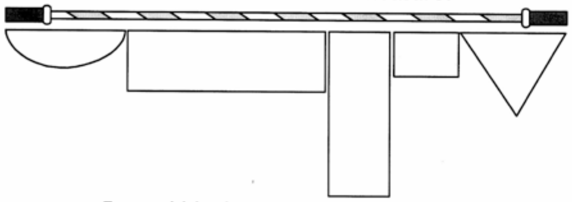
____

What tool would you use to find this information?

a. A graduated cylinder.
b. A measuring tape.
c. A hammer.
d. A scale.

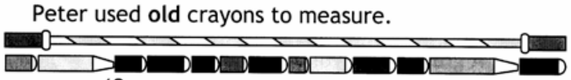
3. Peter, David and Mike used different ways to measure the length of a jump rope.

David used **new** wood blocks to measure.



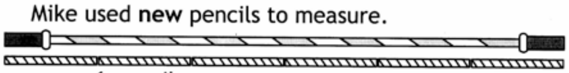
= 5 wood blocks

Peter used **old** crayons to measure.



= 12 crayons

Mike used **new** pencils to measure.



=6 pencils

Which way do you think is the best and why?

____

____

What tool would you use to measure the length of the jump rope?

a. A piece of paper.
b. A string.
c. A scale.
d. A ruler.

Performance Interview

Task 1: **Materials**: a ruler, a graduated cylinder, a beaker, a balance scale.

a) Have you ever seen these things? Do you know what these are called? (if can't come up with names: Do you know what you might use these for?)

b) Could you estimate how long this book is in centimeters? (Is it closer to 2cm, 20cm, 200cm or 1m? Do you know anything that is 2cm, 20cm, 200cm, or 1m?)

c) Could you measure it and tell me how long it is?
   - ✓ select a ruler to measure.
   - ✓ place ruler correctly (align to zero, know to start from 0 again when the book is longer than the ruler)
   - ✓ interpret the result correctly (add up the numbers)

d) My friend, Susan, also has a book. She thinks her book is bigger than mine, but I think my book is bigger than hers. I can ask her questions to figure out whose book is bigger. What kind of questions can I ask her to know whose book is bigger? (if ask irrelevant information: Does that help me know whose book is bigger?) What can she use to measure it? Is there any other way to measure this book?
   - ✓ Suggest that length, width, thickness is the property of interest.
   - ✓ Select a ruler to measure

e) Susan told me she does not have [std tool suggested by child] to measure her book, but she can use other things to measure, such as crayons, AA batteries, and screws. I found these things in my house, but I am not sure if Susan got the same things as mine. What can Susan use to measure her book so that I would know whose book is bigger? (If child takes too long to decide which one to use, then go to <u>can't decide</u>.)

<u>Batteries</u>:

I. Why do you think she should use batteries? Why can't she use crayons or screws to measure? Is there any reason that you think she should use batteries?
   - ✓ Select tool with universal units

II. Could you show me how to use battery to measure the book?
   - ✓ Measure the length of a book using non-standard tool.

III. Susan told me her book is about [A: 7 and a half; B: 6] batteries long. Do you think her book is longer than mine? Why do you think so? Is there any other way to figure out which book is bigger?
   - ✓ Identify width, length, thickness to decide how big a book is.

<u>Crayons or Screws</u>:

I. Why do you think she should use [*crayons, screws*]?
   - ✓ Select tool with identical units

II. Could you show me to use [*crayons, screws*] to measure the book? (I notice that you were using the same [*crayon, screw*] to measure the book, is there any reason why you don't use these [*shorter crayons*, *longer screws*]?)
   - ✓ Measure the length of a book using non-standard tool.

III. Susan told me her book is about [A: 4 and a half; B: 3 and a half] *crayons* or [A: *screws*] long. Do you think her book is longer than mine? Why do you think so? Is there any other way to figure out which is bigger?
   - ✓ Identify width, length, thickness to decide how big a book is.

<u>Can't decide</u>:

I. You can just pick one and we can start from there. (After the child picked one) Could you use [battery, crayon, or screw] to measure this book? How long is this book?
   - ✓ Select tool with universal or identical units.
   - ✓ Measure the length of a book using non-standard tool.

II. Susan told me her book is [batteries: A: 7 ½ ; B: 6][crayons: A: 4 ½ ; B: 3 ½]
[screws: A: ; B:].
Do you think Susan's book is longer than mine? Why do you think so? Is there any other way to find out which is bigger?
   - ✓ Identify width, length, thickness to decide how big a book is

Task 2: Materials: dog food, water, two containers, two cups in different sizes.

a) I have two dogs at home. I need to feed them with dog food and water. If one dog gets more food or water, the other will get upset. I don't want them to get upset, so when I separate the dog food into two halves, they have to be the exact same amount. Could you help me separate the dog food in these two containers?
   - ✓ Identify that weight can be used to separate dog food.
   - ✓ Select a balance scale.
   - ✓ Place and adjust equal amount of dog food to each side of the scale.
   - ✓ Interpret that there are equal amount of dog food in two containers.

b) Could you help me separate their water in two containers?
   - ✓ Identify that weight/volume can be used to separate water.
   - ✓ Select a graduated cylinder or a balance scale to measure water.
   - ✓ Read correctly from the graduated cylinder or balance scale.
   - ✓ Interpret that there are equal amount of water in two containers.

Interview question: follow-up questions about children's written assessment.

This is the test that you wrote before. Did you remember a question that three people use different ways to measure a jump rope? Could you explain to me which way is the best and why?

APPENDIX D

INSTRUMENTS USED FOR PHYSICS OF SOUND MODULE

Selected FOSS written assessment:

1. The same sound will seem louder if it travels through which of the following?
   A. Solid materials
   B. Liquids
   C. Air
   D. Solids, liquids, or gases will sound the same

2. If you tie a string around a doorknob, pull on the string slowly, and keep plucking it with your finger as you pull, what happens to the pitch?
   A. As you tighten the string, the pitch gets lower
   B. The pitch stays the same.
   C. As you tighten the string, the pitch gets higher.
   D. You can't tell unless you try it.

3. How does sound travel from one place to another?

____

____

____

Big Idea Written Assessment:

1. Which of the following two objects will sound the same when dropped on the floor?

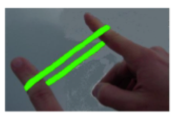
   A. a plastic ball  and a wooden block 

   B. a plastic ball  and a plastic block 

   C. a plastic block  and a wooden block 

   D. they will all sound the same

2. Mary has two **identical** rubber bands except their colors. She stretches one rubber band tighter than the other. What will happen when she plucks the two rubber bands?

   A. The red one will have a lower pitch than the green one.
   B. The green one will have a lower pitch than the red one.
   C. The red one will sound louder than the green one.
   D. The green one will sound louder than the red one.

3. How does a guitar make sound?

____

____

____

Thank You!

Performance Interview:

Task 1:
Ask students to turn around so they can't see what the interviewer is doing. Interviewer turns on a cell phone on the table and asks students:

a. What just happened when you turned around? Where did it happen? How did you know what happened if you didn't see it? How did it get to your ears?
   - Explain sound travels through air.
   - Explain sound is perceived by ears.
b. If I cover this music box with this plastic cup, would you be able to hear anything inside this cup?
   - Predict what will happen if the sound path is blocked.
   - Explain that sound could travel through solids.
c. Put the music box under the plastic cups with strings attached, and ask: I have two sets of plastic cups. They are exactly the same except this set has a string attached to it. If I use these two sets to hear what's inside the cup, which one would sound louder? Why do you think this set would sound louder? Why does sound seem louder through strings?
   - Predict that sound would be louder when sound travels through strings (solids).
   - Explain sound travels better through the string than through the air

Task 2:
Show several xylophone tubes with different lengths to students. Hold the stick so students can't hit the tubes now. Hit one xylophone and ask:

d. Why does it make sound when I hit it? What happened to the xylophone tube when I hit it?
   - energy is applied to make sound.
   - The xylophone tube vibrates when you hit it.
e. I think the sound of this xylophone tube is too low, could you help me pick one with a higher pitch?
   - Select a xylophone tube that makes a higher pitch.
f. ((Wait till students pick one xylophone)). How do you know this one will make a higher pitch than this one? Does that have anything to do with how it vibrates?
   - Explain that the longer tubes make a lower pitch.
   - Explain shorter tube will cause faster vibrations and make higher pitch.
g. Now you can try and see if this one makes a higher pitch than the other one? ((Wait until students tested the pitch)) Does this one make a higher pitch? ((If the student chooses a longer tube and suggests it actually makes a higher pitch after testing, then the student is probably not able to distinguish higher pitch from lower pitch.))
   - Differentiate higher pitch and lower pitch by listening.

Task 3:
Show a set of drums with different tautness. Students can feel the drums before they make predictions.

h. Which one will have a higher pitch? Why do you think it will make a higher pitch? Does that have anything to do with vibrations?
   - Predict the pitch of the drums.
   - Explain the pitch is caused by the tautness of the drums.
i. ((After the student tests the drum.)) Does it make a higher pitch or lower pitch than that one? Why do you think these drums make different pitches?
   - Differentiate high pitch from low pitch.
   - Explain the pitch is caused by the tautness of the drums.

Interview question:

You were doing measurement unit last year. Did you remember how to measure things? I need a table for my apartment, and I want to see if this table could fit in my bedroom. I need to find out how big this table is, right? I didn't bring my ruler so I need to use these things to measure the table. Which one of these things should I use to measure this table? Why? (no need to measure if explained clearly).

- Pick identical unit
- Explain they are all the same size

APPENDIX E

CODING SCHEME AND CODING SHEET OF STUDENT INTERVIEW OF MEASUREMENT

Coding Scheme of Student Talk—Measurement

General: The purpose of this coding scheme is to distinguish students' performance on doing measurement procedures and understandings of measurement concepts. The coding scheme is categorized into procedures and concepts. Each category contains the learning outcomes that were expected to be seen from students' performance.

Measurement Concepts (Causal explanations of measurement):

1[**]. Reasonable estimation: Students were asked to estimate the length of a book to see whether students were able to estimate within a reasonable range of a standard unit**.**

- 0 Not within the range of reasonable value
- 1 Within the range of reasonable value (Reasonable value is set to the actual value ± 15cm) For interviewer A: 20 ~ 50cm; For interviewer B: 15 ~ 45cm
- 9 missing, did not ask

2. Visible strategy: While students were estimating the length of the book, record whether students used any visible gesture to estimate the length.

- 0 no visible strategy
- 1 used some strategy to estimate. *e.g., used pinkie or fingers to make a range and went over the book.*

5. Choice of non-std tools: Students were asked to measure the length of the book with different-sized tools (i.e., blocks, nails) and same-sized tool (i.e., batteries). Record the reasons students provided for choosing a tool.

Uniformity: because they are the same size

- 0 Did not mention anything about uniformity
- 1 Chose similar length crayons or screws while measuring
- 2 Repeated using one crayon or screw to measure
- 3 Mentioned the tools have to be the same size.

[**] The number indicated here corresponded to the numbers of the expected learning outcomes appeared in the coding sheet.

<u>Expedience</u>: because it is longer or because it is faster

0 Did not mention anything about expedience

1 Chose screws or crayons because it is longer or faster

<u>Flush fit</u>: because it fits both ends

0 Did not mention anything about expedience

1 Chose screws or crayons or batteries because it fits end-to-end, or attempted to find something short at the end of the line

Measurement Procedures (Procedural knowledge of measurement):

3. Use of tool – ruler: Students were asked to measure a book with a ruler, which is shorter than the book. Record whether students marked the end of the ruler and iterated the use of the ruler. Students might not see the need to iterate if the difference was considered negligible.

0 Noticed the difference but did not iterate using a ruler

1 Noticed the difference and iterated using a ruler

9 Did not see the need; ignored the difference

4. Identify Property of Interest: Students were asked to compare the book at hand with someone's book which is far away and decided which book is bigger. Record whether students suggested a proper property to measure (i.e., length or width).

0 Was not able to identify a property to measure

1 Suggested to use length or width to measure

2 Suggested more than one property to measure

6. Quantify Dog Food: Students were asked to separate dog food into equal amounts.

<u>First strategy</u>: Record the first attempt students made to separate dog food.

0 Counted the pieces of dog food

1 Visualized the level of dog food in two bowls, or scooped the same number of dog food in each bowl.

2 Used a tool

<u>Use of tool</u>: Record whether students used a tool to separate the dog food.

0 Did not use any tool throughout the task; or being specifically prompted which tool to use

1 Used a tool after prompting (e.g., is there anything on this table that can help me separate the dog food?)

2 Used a tool without prompting

<u>Tool chosen</u>: Record what tool was used to measure

A Balanced a scale, including counting pieces then putting into the scale

B Used a measuring cup then read from it

<u>Ability to use the selected tool</u>: Record whether students were able to use the tool

0 Did not use the tool properly

1 Used the tool properly

7. Quantify water: Students were asked to separate water into equal amounts.

<u>First strategy</u>: Record the first attempt students made to separate water

0 Scooped water into two bowls then checked the water level visually in both bowls

1 Used a tool

<u>Use a tool</u>: Record whether students used a tool to separate water

0 Did not use any tool throughout the task

1 Used a tool after prompting

2 Used a tool without prompting

<u>Tool chosen</u>: Record what tool was used to measure. Children might use more than one strategy to make sure water is the same amount in both bowls. Record the <u>last</u> tool chosen to measure the water.

A Scooped the same number of cups

B Balanced the scale

C Read from the measuring cup

<u>Ability to use the selected tool</u>: Record whether students were able to use the selected tool appropriately

0 Did not use the tool properly

1 Used the tool properly

Subj: ____ Interviewer: __________ Coded by: _____ Today's Date: _________

Expected learning outcomes		Code
1. Reasonable estimation	0=No, 1=Yes	
2. Visible strategy	0=No, 1=Yes	
3. Use of tool: Iterate a ruler	0=mentioned to use a longer ruler 1=no clear mark of the end (flipped or moved the ruler) 2=mark ends 9=did not see the need to iterate	
4. Identify property of interest	0=not sure what to identify 1=identify length or width 2=identify more than one property	
5. Choice of non-std tools	*Uniformity*: 0=did not mention uniformity 1= used similar-sized tool; or used something close to 1cm (pinkie) as tools 2= repeated using one crayon or screw to measure 3=used batteries to measure and mentioned they are the same size	
	Expedience: 0=did not mention expedience 1=used screws to measure and mentioned that it's longer	
	Flush fit: 0=did not mention flush fit 1=explained that because the tool fits end-to-end; or used something short to fit the end of the line	
6. Quantify dog food	*First strategy*: 0=counted the pieces of dog food 1=eyeballed the level of dog food; put same scoops in each bowl 2=used a tool	
	Use of tool: 0=did not use any tool; or being specifically prompted which tool to use 1=used a tool after prompting 2=used a tool without prompting	
	Tool chosen: A=balanced a scale B=used a measuring cup	
	Ability to use tools: 0=did not use the tool properly 1=used the tool properly	
7. Quantify water	*First strategy*: 0=eyeballed the water level in two bowls 1=used a tool	
	Use a tool: 0=did not use any tool 1=used a tool after prompting 2=used a tool without prompting	
	Tool chosen: A=put the same scoops into each cups B=balanced the scale C=read from the measuring cup	
	Ability to use the tool: 0=did not use the tool properly 1=used the tool properly	

APPENDIX F

CODING SCHEME AND CODING SHEET OF THE INTERVIEW OF THE SOUND UNIT

Coding Scheme of Student Talk—Sound

General: The focus of this study is to document students' explanations of sound propagation and properties. Thus, the elaboration of sound path or sound production in students' explanation is given more scores in this rubric.

Understandings of Sound Propagation (path):

Task 1: How does sound travel from the cell phone to your ears?
Phone → ears:

1. did not mention anything about sound path: *e.g. because your cell phone rang and I heard it with my ears.*
2. mentioned a little about how sound travels:
    a. sound travels in a straight line (mention that sound goes directly to your ears): *e.g. sound travels direct to your ears; sound came to my ears.*
    b. sound travels in all directions: *e.g. sound spreads everywhere*
3. mentioned how sound travels (through medium or because of vibrations):
    a. sound needs a medium to travel (mention what happened between sound source and sound receiver): *e.g. sound travels through the air (whereas "the air is inside the cup, since the air is trapped, they can't get the sound out." This statement should be considered as "2" because air was used as a replacement for "sound" instead of seen as "medium")*
    b. sound travels because of vibrations (mention specifically that sound travels through vibrations): *e.g. sound vibrates through the air*

Task 2: If I cover this cell phone with this cup, will you be able to hear sound?
Phone → cup → ears:

1. did not mention microscopically about sound path, such as referring to past experience of an alarm clock or someone's cell phone: *e.g. sound is loud so you can hear it.*
2. mentioned about the sound path but did not mention correctly how sound is transmitted (through vibrations):
    a. sound path being blocked: *e.g. sound will be blocked by the cup*
    b. sound spreads out from little holes: *e.g. sound can get out from the bottom of the cup*
3. mentioned that sound travels by vibrating: *e.g. sound inside the cup was vibrating and the cup was vibrating too.*

Task 3: Two sets of cups: one set is attached with a string and the other set without a string. Which one will you be able to hear better sound?
Phone → string → ears:

1. did not mention path or vibration, despite of knowing that string cups will be louder: *e.g., because it's loud.*
2. mentioned path that sound travels through a medium, such as through the string or through the hole in the cup, but without further explanation about sound transmission: *e.g. sound will go through the string, but it will be quieter. Sound will get out from the bottom of the cup, so it will be louder.*
3. mentioned sound path and knew that sound will be louder using a string, but no further explanation about sound transmission: *e.g. it might go through the air (gestures somewhere outside the string) and I might be able to hear it when I put my ear on.*
4. mentioned sound path and vibration but no indication of knowing that sound is louder through solid materials: *e.g. sound can go through the string whereas the other set of cups, sound will spread out in the air*. *e.g. sound goes straight through the string to your ear instead of everywhere else*
5. mentioned that sound travels through vibrations and is louder through the string: *e.g. the sound of the phone makes the string vibrates and then comes to my ears.*

Understandings of Sound Production:

Task 1a: explain why sound is made when I hit the xylophone tube.
Xylophone tube:

1. mentioned irrelevant factor of sound,
    a. sound is made because of its material: e.g. *because it's metal*
2. mentioned how sound is produced by referring to action,
    a. sound is made by personal action: *e.g. because you hit it*
3. mentioned sound is made because of applying energy or the vibrations of an object: *e.g. because you put energy into it; because the tube vibrates*

Task 1b: Which tube makes a higher/lower pitch sound and why?
Pitch vs. properties of the object

1. no indication of relevant factor or did not mention any factor: *e.g. this one has a higher pitch because it is hard*.
2. mentioned length as a factor, but in wrong direction: *e.g., because it's big and it spreads all around this room.*
3. mentioned length as a factor and in the right direction, but without explanations: *e.g. this one has a higher pitch because it is shorter.*
4. mentioned length as a factor and attempted to provide explanations about making different pitch sound: *e.g. it is shorter and the air can go through it faster; it is shorter and it vibrates faster.*

Task 2: Which drum makes a higher pitch sound and why?
Drums (pitch vs. properties of the drum):

1. did not mention any relevant factor: *e.g. this one has a higher pitch because this purple line is lower than the other ones*
2. mentioned something relates to tautness as a factor: *e.g., this one is thinner so it makes the highest pitch; this one has a bigger bump.*
3. mentioned tautness as a factor with (code 4) or without explanations:
    a. *e.g. this one has a lower pitch because it is damp.*
    b. Recall from past experiences: *e.g. I know this because my mom made drums like this sometimes*.
    c. *e.g. because the drum is tight and tighter things vibrates faster to make higher sound.*

Drums (observe high and low pitches after testing)

1. did not mention pitches in the right direction: *e.g., the loosest drum has the highest pitch.*
2. mentioned pitches in the right direction.

Subj: ____ Interviewer: __________ Coded by: _____ Today's Date: _________

		Explanations of sound path or production	**Students' descriptions of sound**
Sound propagation	Phone → ears	0. no sound path 1. mentioned path but no vibrations or medium 2. mentioned vibrations or medium	
	Phone→cup →ears	0. no sound path 1. mentioned path but no vibrations or medium 2. mentioned vibrations or medium	
	Phone→string →ears	0. did not mention path or vibration 1. mentioned path but don't know string will be louder 2. mentioned path and knows string is louder 3. mentioned path and vibrations but don't know string is louder 4. mentioned string is louder because of vibrations	
Sound Production	Xylophone tube	0. referring to material 1. referring to personal action 2. apply energy or vibrations	
	Pitch vs. length	0. irrelevant factor 1. mentioned length but in wrong direction 2. mentioned length but no explanation 3. mentioned length with attempts to provide explanations	
	Pitch vs. tautness	0. irrelevant factor 1. mentioned something relates to tautness but no explanations 2. mentioned tautness with or without attempts to provide explanations	
	Differentiate pitch	0. not able to tell 1. able to tell which pitch is higher	
	Measurement	1. wood block 2. batteries 3. lego blocks	Reason: same size, longer, other

APPENDIX G

CODING SCHEME AND CODING SHEET OF CLASSROOM OBSERVATIONS

Coding Scheme of Teacher Talk

(Revised from Newton & Newton, 2000[††])

This coding system is aimed to identify the type of inquiry opportunity that teacher provided to students while enacting hands-on science curriculum in the classroom. Such inquiry opportunities can be shown from what teacher asked, said, or provided to students. The focus of this coding scheme is to identify what kind of thinking students were engaged in through the questions teacher asked or requested, is it mostly about factual knowledge, vocabulary to be memorized, describing observations, explaining what happened, or following procedures.

(1) Identify topic related set (TRS): Read through the field note and identify different topics of the lesson, such as review of previous lessons, introduction of group activities, demonstration of hands-on activity, summary of hands-on activity, etc. Each topic can be identified by switching to different forms of activities or teacher's introduction of another task.

(2) Locate questions or requests that teacher asked: Find out every question that teacher <u>initiated</u> which related to the science topic at hand and which required students' responses. In the situations that teacher repeats, revoices or rephrases students' answers those repetitive questions should be clustered as one question. Use student response to help decide whether different teacher questions should be clustered and what type of inquiry opportunities was intended. If students provided different concept on the same question, or if teacher elaborate on each student response regarding the same question, do not cluster repetitive questions as one question.

[††] Newton, D. P. & Newton, L. D. (2000). Do teachers support causal understanding through their discourse when teaching primary science? *British Educational Research Journal, 26(5),* 599-613.

(3) **Identify types of inquiry opportunities.** Categorize each unit as DF, CE, Proc, or Others according to the following descriptions.

a. **DF (descriptive/factual knowledge):** aimed to elicit descriptions of a phenomena or factual response of science content.
b. **CE (causal explanatory knowledge):** aimed to elicit patterns or explanations of phenomena
c. **Proc (procedural knowledge):** aimed to elicit proper steps or directions of doing science activities; mostly about next actions. The elicited procedures include those established by teacher in previous lessons, and those provided by students, such as teacher asks students to find out what to do to solve a particular problem.
d. **Others:** things that can't be categorized as DF/CE/Proc but still relate to the science content.

Code	Descriptions	Examples
Descriptive and Factual (DF): Aimed to elicit descriptions of a phenomena or factual response of science content		
Code	Descriptions	Examples
Fact	aimed to elicit a factual response such as naming, stating attributes, times, locations	*Which one is sound source? What is the standard unit of length?*
Vocab	aimed to introduce or recall a concept of vocabulary	*What does amplify mean?*
Sum	aimed to summarize what was observed or what was done in the previous activities	*What did we find out yesterday?*
Dscp	aimed to elicit the descriptions of a phenomena, situation or event	*What did this look like? What happened?*
Causal Explanations (CE): Aimed to elicit patterns or explanations of phenomena		
Pattn	aimed to find or recall a pattern or relation from a phenomenon, situation or event	*What did we know about long and short strings? Long string makes a low pitch and short string makes a high pitch.*
Exp	aimed to elicit an explanation of a situation or event involving physical causation	*Why were you able to hear sound when you're not there? Why does the shorter string make a higher pitch?*
Prdct	aimed to make predictions about a situation or event with established explanations before	*How do we make high pitch? This will have a higher pitch sound because it's shorter.*
Procedures of science activities (Proc): Aimed to elicit proper steps or directions of doing science activities; mostly about next actions Procedures can be seen as "actions which can be done through step-by-step instructions"		
Proc	Remind students the steps related to the ongoing or past activity	*What should I do after this? Do you understand what I meant by doing this and that?*
Cal	Record results or organize results in a graph or table	*I have 50ml, 50ml, 28ml, what do I have in total? I have 25 Celsius, where should I plot this?*
Prgrss	Elicit the progress of an ongoing activity	*Have you finished this? Did you do this?*
Others (O):		
Other	Evaluate or reflect on an activity	*What was easy or difficult in today's activity?*
	Provide rationale for solutions or actions	*How do we prove that they get the same amount of lemonade.*

APPENDIX H

POST HOC TESTS OF THE MEASUREMENT MODULE

Codes	Student learning outcomes	Percentage of correct response			Post hoc tests (P-value) (using Mann-Whitney U)		
Causal Explanatory (CE)	**-Making estimations with standard units**	Sarah	Clark	Margaret	Sarah vs. Clark	Clark vs. Margaret.	Margaret vs. Sarah
	Estimate length of a piece of paper (WA:FOSS_3)	92%	25%	73%	.010*	.071*	
	Estimate mass of a piece of paper (WA:FOSS_11)	15%	25%	60%			.046*
	Estimate length of a book (PI:1b)	85%	45%	93%		.041*	

APPENDIX I

POST HOC TESTS OF THE SOUND MODULE

Codes	Student learning outcomes	Percentage of correct response or average score			Post hoc tests (P-value) (using Mann-Whitney U)		
	-Properties of sound	Sarah	Clark	Margaret	Sarah vs. Clark	Clark vs. Margaret.	Margaret vs. Sarah
Descriptive/ Factual (DF)	Same materials make similar sounds (WA†:BI_Q1)	100%	18%	50%	.009**		
	Identifying length as a factor to sound pitch of tubes (PI‡:#2)	100%	67%	100%	.032*	.022*	
	Tighter rubber band makes a higher sound (WA:BI_Q2)	80%	9%	43%	.027*		

Codes	Student learning outcomes	Percentage of correct response or average score			Post hoc tests (P-value) (using Mann-Whitney U)		
Causal Explanatory (CE)	**-Sound production**						
	Explain how guitar made sounds. (WA†:BI_Q3)	Mean:1.00 (SD: 0)	0.82 (0.405)	1.36 (0.497)		.044*	
	Explain how xylophone tubes made sound. (PI‡:#2)	1.23 (0.599)	1.00 (0.632)	1.80 (0.414)		.018*	.022*
	Explain why shorter tubes made higher pitches.(PI:#2)	0.23 (0.438)	0 (0)	0.60 (0.507)		.036*	
	-Sound transmission						
	Explain how sound travels (WA:FOSS_Q3)	1.40 (0.894)	0.45 (0.522)	1.57 (0.646)		.005**	
	Sound is louder when travels through solids (FOSS: Q1)	40%	9%	57%		.044*	

APPENDIX J

SYNTHESIZED RESULTS OF STUDENT LEARNING OUTCOMES AND TEACHER TALK

The measurement module

	Student learning outcomes			Proportions of teacher talk		
Codes		Sig. results	Insig. results	Sarah	Clark	Margaret
Descriptive/ Factual (DF)				55%	32%	51%
Causal Explanatory (CE)	**-Making estimations with standard units**			0%	2%	1%
	Estimate length of a piece of paper (WA:FOSS_3)	.005*				
	Estimate height of an adult (WA:FOSS_15)		.801			
	Estimate mass of a piece of paper (WA:FOSS_11)	.044*				
	Estimate length of a book (PI:1b)	.012*				
	-Using identical units to measure					
	Identify measurement tools with identical units (WA:BI_3)		.122			
	Identify measurement tools with identical units (PI:1e)		.157			
Procedural (Proc)	**-Identify property of interest:**			26%	44%	41%
	Measure the space in between (WA: BI_2a)		.698			
	Compare the size of books (PI: 1d)		.571			
	-Select appropriate tools:					
	Choose thermometer to measure temperature (WA: FOSS_4)		.427			
	Choose rulers to measure length (WA:BI_2c)		.258			
	Choose rulers to measure length (WA:BI_3b)		.277			
	Choose rulers to measure length (PI: 1c,1d)		---			
	Use tools to separate dog food (PI: 2a): with prompting without prompting		.244 .234			
	Use tools to separate water (PI: 2b): with prompting without prompting		.417 .285			
	-Place materials appropriately:					
	Draw marbles to make a scale balanced (WA: BI_1b)		.316			
	Measure length of a book (PI:1c)		---			
	Separate dog food and water in half (PI: 2a)		---			

	-Report results properly:					
	Report results off a balance scale (WA:FOSS_14)		.690			
	Report results off a thermometer (WA:FOSS_24a)		.271			
	Report results off a thermometer (WA:FOSS_24b)		.352			
	Report results off a graduated cylinder (WA:FOSS_26)		.614			
	Predict results of a scale (WA:BI_1a)		.191			
	Balance a scale (PI:2a)		---			
	Read a measuring cup (PI:2b)		---			

The sound module

	Student learning outcomes			Proportions of teacher talk		
Codes		Sig. results	Insig. results	Sarah	Clark	Margaret
Descriptive/ Factual (DF)	**-Properties of sound**			61%		71%
	Same materials make similar sounds (WA†:BI_Q1)	.009**				
	Identifying length as a factor to sound pitch of tubes (PI‡:#2)	.007**				
	Tighter rubber band makes a higher sound (WA:BI_Q2)	.019*				
	Tighter string makes a higher sound (WA:FOSS_Q2)		.355			
Causal Explanatory (CE)	**-Sound production**			19%		10%
	Explain how guitar made sounds. (WA†:BI_Q3)	.018*				
	Explain how xylophone tubes made sound. (PI‡:#2)	.006**				
	Explain why shorter tubes made higher pitches.(PI:#2)	.019*				
	Explain why tighter drums made higher pitches (PI:#3)		.253			
	-Sound transmission					
	Explain how sound travels (WA:FOSS_Q3)	.002*				
	Explain how sound travels to your ears (PI:#1)		.150			
	Explain how sound travels when covered with a cup (PI:#1)		.457			
	Sound is louder when travels through solids (FOSS: Q1)	.046*				
	Comparing sound travels with or without a string (PI:#1)		.816			
Procedural (Proc)				14%		19%

BIBLIOGRAPHY

American Association for the Advancement of Science (1993). Benchmarks for Science Literacy. New York: Oxford University Press.

Anderson, R. D. (2002). Reforming science teaching: What research says about inquiry. Journal of Science Teacher Education, 13, 1-12.

Anderson, R. D. & Helms, J. V. (2001). The ideal of standards and the reality of schools: Needed research. Journal of Research in Science Teaching, 38, 3-16.

Ball, D. L. (1994). Developing mathematics reform: What don't we know about teacher learning—but would make good working hypotheses? Paper presented at Conference on Teacher Enhancement in Mathematics K-6, Arlington, VA.

Ball, D. L., & Cohen, D. K. (1996). Reform by the book: What is—or might be—the role of curriculum materials in teacher learning and instructional reform? Educational Researcher, 25(9), 6-8, 14.

Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In G. Sykes and L. Darling-Hammond (Eds.), Teaching as the learning profession: Handbook of policy and practice (pp. 3-32). San Francisco: Jossey Bass.

Barrett, J. E. & Clements, D. H. (2003). Quantifying path length: Fourth-grade children's developing abstractions for linear measurement. Cognition and Instruction, 21, 475-520.

Baxter, G. P., Shavelson, R. J., Goldman, S. R., & Pine, J. (1992). Evaluation of a procedure-based scoring for hands-on science assessment. Journal of Educational Measurement, 29(1), 1-17.

Ben-Peretz, M. (1990). The teacher-curriculum encounter: Freeing teachers from the tyranny of texts. Albany, NY: State University of New York Press.

Borko, H. & Putnam, R.T. (1996). Learning to teach. In R.C. Calfee & D.C. Berliner (Eds.), Handbook of educational psychology. New York: Macmillan.

Bredderman, T. (1983). Effect of activity-based elementary science on student outcomes: A quantitative synthesis. Review of Educational Research, 53, 499-518.

Bybee, R. W. and Fuchs, B. (2006) Preparing the 21st-century workforce: a new reform in science and technology education. Journal of Research in Science Teaching, 43(4) 349–352.

Carter, L. (2005). Globalisation and science education: Rethinking science education reforms. Journal of Research in Science Teaching, 42, 561-580.

Cartier, J. L. (2003). Implementing inquiry pedagogy in elementary schools. National Science Foundation Research Protocol, NSF ADVANCE Award #0137730.

Cartier, J. L., Barton, A. M., & Mesmer, K. (2001). Learning through and about inquiry in a 9th-grade earthbased astronomy unit. Paper presented at the annual meeting of the National Association for Research in Science Teaching, St. Louis, MO, March 2001.

Cartier, J. L., Passmore, C. M., and Stewart, J. (2001). Balancing generality and authenticity: A framework for science inquiry in education. Paper presented at the annual meeting of the International History, Philosophy, and Science Teaching Organization, Denver, CO, November 7-10.

Cartier, J. L., Passmore, C. M., Stewart, J., & Willauer, J. P. (2005). Involving students in realistic scientific practice: Strategies for laying epistemological groundwork. In R. Nemirovsky, A. S. Roseberry, J. Soloman, & B. Warren (Eds.), Everyday matters in science and mathematics (pp. 267-298). Mahwah, NJ: Lawrence Erlbaum Associates.

Cawley, J. F. & Cawley, L. J. (1994). Science education for students with disabilities. Remedial and Special Education, 15, 67-133.

Clandinin, D. J., & Connelly, F. M. (1992). Teacher as curriculum maker. In P. W. Jackson (Ed.), Handbook of research on curriculum (pp. 363-410). New York: Macmillan.

Clements, D. H. (1999a). Geometric and spatial thinking in young children. In C. Juanita (Ed), Mathematics in the Early Years. Washington, DC, US: National Association for the Education of Young Children, 66-79.

Clements, D. H. (1999b). Teaching length measurement: Research challenges. School Science and Mathematics, 99, 5-11.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. Educational Researcher, 32, 9-13.

Crawford, B. A. (2000). Embracing the essence of inquiry: New roles for science teachers. Journal of Research in Science Teaching, 37, 916-937.

Dalton, B., Morocco, C. C., Tivnan, T., & Mead, P. L. R. (1997). Supported inquiry science: teaching for conceptual change in urban and suburban science classrooms. Journal of Learning Disabilities, 30, 670-684.

Davis, E. A., & Krajcik, J. S. (2005). Designing educative curriculum materials to promote teacher learning. Educational Researcher, 34(3), 3-14.

DeBoer, G. E. (2004). Historical perspectives on inquiry teaching in schools. In L.B. Flick & N.G. Lederman (Eds.), Scientific inquiry and nature of science: Implications for teaching, learning, and teacher education (pp. 17-35). Dordrecht, The Netherlands: Kluwer.

Driver, R., Squires, A., Rushworth, P., & Wood-Robinson, V. (1994). Making sense of secondary science: Research into children's ideas. New York: Routledge.

Eshach, H., & Schwartz, J. L. (2006). Sound stuff? Naïve materialism in middle-school students' conceptions of sound. International Journal of Science Education, 28(7), 733-764.

Greeno, J. G. (1997). On claims that answer the wrong questions. Educational Researcher, 26(1), 5-17.

Hapgood, S., Magnusson, S. J., & Palincsar, A. S. (2004). Teacher, text, and experience: A case of young children's scientific inquiry. The Journal of the Learning Sciences, 13(4), 455-505.

Hiebert, J. (1984). Why do some children have trouble learning measurement concepts? Arithmetic Teacher, 31, 19-24.

Hofstein, A. & Lunetta, V. N. (1982). The role of the laboratory in science teaching: Neglected aspects of research. Review of Educational Research, 52, 201-217.

Huber, R. A. & Moore, C. J. (2001). A model for extending hands-on science to be inquiry-based. School Science and Mathematics, 101, 32-42.

Joram, E., Subrahmanyam, K., & Gelman, R. (1998). Measurement estimation: Learning to map the route from number to quantity and back. Review of Educational Research, 68, 413-449.

Kamii, C., & Clark, F. (1997). Measurement of length: The need for a better approach to teaching. School Science and Mathematics, 97, 116-121.

Lautrey, J., & Mazens, K. (2004). Is children's naïve knowledge consistent? A comparison of the concepts of sound and heat. Learning and Instruction, 14, 399-423.

Lave, J., (1988). Cognition in practice: Mind, mathematics, and culture in everyday life. Cambridge: Cambridge University Press.

Lee, O., Buxton, C., Lewis, S., & LeRoy, K. (2006). Science inquiry and student diversity: Enhanced abilities and continuing difficulties after an instructional intervention. Journal of Research in Science Teaching, 43, 607-636.

Lehrer, R. (2003). Developing understanding of measurement. In J. Kilpatrick, G. Martin, & D. Schifter (Eds.). Research Companion to Principles and Standards for School Mathematics (pp. 179-192), Reston, VA: National Council of Teachers of Mathematics.

Lehrer, R., & Schauble, L. (2002). Investigating real data in the classroom: Expanding children's understanding of math and science. New York: Teachers' College Press.

Linder, C. J. (1993). University physics students' conceptualizations of factors affecting the speed of sound propagation. International Journal of Science Education, 15, 655-662.

Linder, C. J., & Erickson, G. L. (1989). A study of tertiary physics students' conceptualizations of sound. International Journal of Science Education, 11, 491-501.

Lynch, S. J. (2000). Equity and science education reform. Mahwah, NJ: Lawrence Erlbaum Associates.

Marx, R. W., Blumenfeld, P. C., Krajcik, J. S., Fishman, B. Soloway, E., Geier, R., Tali Tal, R. (2004). Inquiry-based science in the middle grades: Assessment of learning in urban systemic reform. Journal of Research in Science Teaching, 41, 1063-1080.

Mastropieri, M. A., & Scruggs, T. E. (1994). Text versus hands-on science curriculum. Remedial and Special Education, 15, 72-85.

Mazens, K., & Lautrey, J. (2003). Conceptual change in physics: Children's naïve representations of sound. Cognitive Development, 18, 159-176.

McCarthy, C. B. (2005). Effects of thematic-based, hands-on science teaching versus a textbook approach for students with disabilities. Journal of research in science teaching, 42, 245-263.

McLaughlin, M., & Talbert, J. E. (1993). Contexts that matter for teaching and learning: Strategic opportunities for meeting the nation's educational goals. Stanford, CA: Center for Research on the Context of Secondary School Teaching, Stanford University.

Mehan, H. (1979). Learning lessons: Social organization in the classroom. Cambridge, MA: Harvard University Press.

Mehan, H. (1982). The structure of classroom events and their consequences for student performance. In P. Gilmore & A.A. Glatthornde (Eds.), Children in and out of school. Washington, DC: Center for Applied Linguistics.

National Research Council. (1996). National Science Education Standards. Washington DC: National Academy Press.

Newton, D. P. & Newton, L. D. (2000). Do teachers support causal understanding through their discourse when teaching primary science? British Educational Research Journal, 26, 599-613.

Niemi, D. (1996). Assessing conceptual understanding in mathematics: Representations, problem solutions, justifications, and explanations. The Journal of Educational Research, 89, 351-363.

Piaget, J. (1971). Les explications causales. Paris: PUF.

Piaget, J., Inhelder, B., & Szeminska, A. (1981). The child's conception of geometry (E. A. Lunzer, Trans.), New York: Norton. (Original work published 1960).

Pine, J., Aschbacher, P., Roth, E., Jones, M., McPhee, C., Martin, C., Phelps, S., Kyle, T., & Foley, B. (2006). Fifth graders' science inquiry abilities: A comparative study of students in hands-on and textbook curricula. Journal of Research in Science Teaching, 43, 467-484.

Porter, A. C., & Smithson, J. L. (2001). Defining, developing, and using curriculum indicators. CPRE Research Report Series. Philadelphia: Consortium for Policy Research in Education.

Putnam, R. T., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? Educational Researcher, 29(1), 4-15.

Raghavan, K., Cohen-Regev, S., & Strobel, S. A. (2001). Student outcomes in a local systemic change project. School Science and Mathematics, 101(8), 417-426.

Reiner, M., Slotta, J. D., Chi, M. T. H., & Resnick, L. B. (2000). Naïve physics reasoning: A commitment to substance-based conceptions. Cognition and Instruction, 18(1), 1-34.

Remillard, J. T. (1999). Curriculum materials in mathematics education reform: A framework for examining teachers' curriculum development. *Curriculum Inquiry, 29*(3), 315-342.

Remillard, J. T. (2000). Can curriculum materials support teachers' learning? *Elementary School Journal, 100*(4), 331-350.

Remillard, J. T. (2005). Examining key concepts in research on teachers' use of mathematics curricula. Review of Educational Research, 75, 211-246.

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. Journal of Educational Psychology, 93, 346-362.

Rogoff, B. (1994). Developing understanding of the idea of communities of learners. Mind, Culture, and Activity, 4, 209-229.

Ruby, A. (2001). Hands-on science and student achievement. Unpublished doctoral dissertation, RAND Graduate University.

Salmon, W. (1998). Causality and explanation. New York: Oxford University Press.

Schauble, L., & Glaser, R. (1990). Scientific thinking in children and adults. In D. Kuhn (Ed.), Developmental Perspectives on Teaching and Learning Thinking Skills, 21, 9-27.

Schneider, R. M., Krajcik, J., & Blumenfeld, P. (2005). Enacting reform-based science materials: The range of teacher enactments in reform classrooms. Journal of Research in Science Teaching, 42, 283–312.

Scruggs, T. E. & Mastropieri, M. M. (1994). The construction of scientific knowledge by students with mild disabilities. The Journal of Special Education, 28, 307-321.

Shavelson, R., Baxter, G. P., & Pine, J. (1991). Performance assessment in science. Applied Measurement in Science, 4(4), 347-362.

Shavelson, R. J., Baxter, G. P., & Pine, J. (1992). Performance assessments: Political rhetoric and measurement reality. Educational Researcher, 21, 22-27.

Sherin, B., Edelson, D., & Brown, M. (2004). In L.B. Flick & N.G. Lederman (Eds.), Scientific inquiry and nature of science: Implications for teaching, learning, and teacher education (pp. 221-248). Dordrecht, The Netherlands: Kluwer.

Shymansky, J. A. (1989). What research says...About ESS, SCIS, and SAPA. Science and Children 26: 33–35.

Solano-Flores, G., & Shavelson, R. J. (1997). Development of performance assessments in science: Conceptual, practical and logistical issues. Educational Measurement, Issues and Practice, 16-25.

Stephan, M., & Clements, D. H. (2003). Linear, area and time measurement in prekindergarten to grade 2. In D. H. Clements & G. Bright (Eds.), Learning and teaching measurement (Vol. 65, pp. 3-16). Reston, VA: National Council of Teachers of Mathematics.

Von Secker, C. E. & Lissitz, R. W. (1999). Estimating the impact of instructional practices on student achievement in science. Journal of Research in Science Teaching, 36, 1110-1126.

Warwick, P. & Stephenson, P. (2002). Reconstructing science in education: Insights and strategies for making it more meaningful. Cambridge Journal of Education, 32 (2) 143-151.

Wilkening, F. & Huber, S. (2002). Children’s intuitive physics. In U. Goswami (Ed.), Blackwell Handbook of Childhood Cognitive Development (pp. 349-370). Oxford: Blackwell.