TENNESSEE CONSORTIUM ON RESEARCH, EVALUATION & DEVELOPMENT

THEC STEM Professional Development Program: Final Evaluation Report

JUNE 2014

CARLA C. JOHNSON

VANDERBILT PEABODY COLLEGE

The Tennessee Consortium on Research, Evaluation, and Development is a research and policy center at Peabody College of Vanderbilt University's Peabody College. The Consortium is funded through the State of Tennessee's Race to the Top grant from the United States Department of Education (grant #S395A100032).

Please visit www.tnconsortium.org to learn more about our program of research and recent publications.

This study was supported by the Tennessee Consortium on Research, Evaluation, and Development (the Consortium) at Vanderbilt University's Peabody College, which is funded by the State of Tennessee's Race to the Top grant from the United States Department of Education (grant #S395A100032). Any errors remain the sole responsibility of the author. The views expressed in this paper do not necessarily reflect those of sponsoring agencies or individuals acknowledged.

Direct all correspondence to Carla C. Johnson, Purdue University, 100 N. University Street, West Lafayette, Indiana 47907 (carlacjohnson@purdue.edu).

Tennessee Consortium
on Research, Evaluation, & Development

Peabody #44 | 230 Appleton Place | Nashville, Tennessee 37203
Phone 615-322-5538 | Fax 615-322-6018
www.tnconsortium.org

TENNESSEE CONSORTIUM ON RESEARCH, EVALUATION & DEVELOPMENT

VANDERBILT PEABODY COLLEGE

TABLE OF CONTENTS

SECTION FOUR

SECTION FIVE

List of Figures

List of Tables

EXECUTIVE SUMMARY

In 2010, as part of Tennessee's Race to the Top grant, The Tennessee Higher Education Commission (THEC) received funding for the implementation of STEM (science, technology, engineering, and mathematics) Professional Development (STEM PD) across the state of Tennessee. A request for proposals (RFP) was released in April 2011 for Round One Programs, then again in spring 2012 for Round Two Programs. This RFP focused on delivery of professional development designed to promote innovative practices in STEM education, and participating programs were expected to emphasize the improvement of STEM teacher pedagogical skills and content knowledge. Round One (2011-12) funding was distributed across 11 programs, and 18 programs were funded in Round Two (2012-13). This report includes the overall analysis of the entire THEC STEM Professional Development portfolio, including all 29 funded programs in Round One and Two. The research questions that guided this evaluation included:

1. What impact, if any, do THEC STEM professional development programs have on teachers' pedagogical skills and STEM content knowledge?
2. What impact, if any, do THEC STEM professional development programs have on teachers' opinions regarding the teaching of STEM?
3. Which funded STEM professional development programs demonstrate significant growth in Teacher Quality (pedagogical skills and content) and should be considered for inclusion as best practice for Tennessee?

Round One of the THEC STEM PD program included:

- Three high school Chemistry focused programs (Tennessee Technological University, *Transforming Matter and Classrooms,* Lipscomb University, *Hands on Chemistry,* and East Tennessee State University*, Modeling Instruction of Chemistry in High School*),
- Two elementary science programs (Tennessee Technological University, *Embedding Inquiry and Technology/Engineering Standards into Physical Science in Grades 3-5,* East Tennessee State University, *Reaching for Excellence in Elementary Science through Inquiry, Standards, and PBL*),
- Four primary/elementary school level mathematics programs (Tennessee Technological University, *Numeracy and Multiple Representations for Grades 1-3,* Austin Peay State University, *Momentum – Building Capacity for Change Through Connections,* Middle Tennessee State University, *EMPOWER,* and University of Tennessee at Chattanooga, *Numeracy*), and
- Two middle school level mathematics programs (Tennessee Technological University, *Developing Middle School Mathematics Teachers Pedagogical Content Knowledge,* University of Tennessee at Chattanooga, *TELMU Technology/Engineering + Literacy = Mathematics Understanding*).

Round Two of the THEC STEM PD program included:

- One high school mathematics focused programs (Lipscomb University, *Functions of Algebra*),
- Two high school science programs (East Tennessee State University, *PCMI,* and Lipscomb University, *Integrating STEM: The Power of Science*),
- Two high school mathematics and science programs (Tennessee Technological University, and Roane State Community College, *Designing the Future*, Middle Tennessee State University, *StaRT,*),
- One middle/high school science programs (University of Tennessee at Chattanooga, *Learning Science through Writing*),
- One middle/high school mathematics and science program (University of Memphis, *mMIND),*
- Five middle school focused mathematics and science programs (Middle Tennessee State University, *Project UC STEM,* University of Memphis, *Professional Development for Grades 5-8,* East Tennessee State University, *Incorporating Active Learning into Life Sciences,* Lipscomb University, *Making Mathematics Matter,* Tennessee Technological University, *From Earth to Space with STEM*)
- One elementary school science mathematics and science program (Tennessee Technological University, *STEM Around Us*),
- Two middle school science program (University of Memphis, *Water, Water, Everywhere,* University of Tennessee at Martin, *Integration for Middle School Teachers*),
- One early childhood science program (East Tennessee State University, *Project SEE*), and
- One early childhood mathematics and science program (East Tennessee State University, *Integrating Hands-on STEM Activities with Math and Reading Common Core Standards*, Tennessee Technological University, *Shaping Early STEM Learning*).

CORE CONCEPTUAL FRAMEWORK

THEC STEM PD programs were required to organize the delivery of their programs around the Core Conceptual Framework for Effective Professional Development (Desmione, 2009) as the organizing framework. The five components of the framework include: content knowledge focus, active learning experiences, coherence with state/district goals and standards, extended duration of program, and collective participation of teams of teachers from individual schools. Funded programs described within their proposals how they would address each of the five components of the framework within the context of their STEM PD.

Study Methods

This evaluation used both qualitative and quantitative data to determine the impact of the THEC STEM PD programs. Data collection included teacher classroom observations (video-recorded), two teacher surveys, and program developed pre/post assessments of mathematics or science content knowledge.

Classroom Observations

Each teacher was required to submit three recordings of their teaching: one prior to participation in the THEC STEM PD program, one mid-way through the program, and the final video at the end of the program. Each video was scored using the Local Systemic Change Classroom Observation Protocol (LSC), which was developed by Horizon Research for use with the National Science Foundation's (NSF) State Systemic Initiatives (SSI) as a measure of reform-based instructional practices in science and mathematics. The instrument examines design of lesson, implementation of lesson, culture of instruction, and content knowledge delivered.

Teacher Surveys

Participants also completed two surveys in a pre/post manner for the THEC STEM PD programs. The first survey was the Local Systemic Change Teacher Questionnaire (LSCTQ) appropriate to their content and grade level (e.g., science or mathematics, K-6 or 7-12). The LSCTQ was also designed for use with NSF's SSI programs. The Survey of Enacted Curriculum (SEC) was the second survey used for the THEC STEM PD programs. The SEC survey was developed by the SEC Collaborative and used extensively to evaluate STEM teaching quality and alignment of instruction to academic standards.

Program-Developed Pre/Post Content Assessments

Each program was required to develop their own 25-item pre/post content knowledge assessment for participating teachers to complete. Programs provided copies of their assessments, keys, and spreadsheets of individual item responses for the evaluation.

KEY FINDINGS

Overall Findings

Classroom Observations

Overall, the THEC STEM PD programs significantly improved in all four domains (design, implementation, culture, and content) from baseline to end of program. (See Table ES2.) Design of lesson includes the planning, organization, resources, attention to equity, level of collaboration, flow of lesson, assessments, and sense making that take place during the delivery of lesson. Implementation of lesson consists of the level of investigative mathematics/science included, quality of management of classroom, pace of lesson, modifications made, questioning strategies, and formative assessments included in the delivery of the lesson. Classroom culture refers to the amount of active participation of all students and level of collaborative learning, including having students explore their own ideas, questions, conjectures, and propositions or to challenge the ideas of others. Finally, the mathematics/science content knowledge domain focuses on the accuracy of content knowledge delivered by the teacher, as well as the alignment of content to appropriate grade and student levels of understanding.

Each item within each domain ranges is scored on a scale of 0 to 5, with 0 being used when there is no evidence of a component within a domain, and a score of 5 awarded when a component is used "to a great extent". Each domain has multiple questions that are scored individually, and an overall rating (i.e., mean score) for each domain is generated (see Table ES1).

Table ES1. LSC Overall Rating

Score	Title
0-1.9	Ineffective Instruction
2-2.9	Elements of Effective Instruction
3-3.9	Beginning of Effective Instruction
4-4.9	Accomplished, Effective Instruction
5	Exemplary Instruction

Table ES2. Classroom Observation Findings – All Programs

Domain	Baseline Rating	End Rating	End Classification
Design	2.39	2.88	Elements of Effective Instruction
Implementation	2.61	3.30	Beginning of Effective Instruction
Classroom Culture	2.84	3.48	Beginning of Effective Instruction
Content Knowledge	2.90	3.50	Beginning of Effective Instruction

Teacher Surveys

Teacher surveys included the following constructs: teacher opinions, teacher perceived importance, instructional influences, teacher preparedness, frequency of use of effective pedagogy, student activities, parental support, principal support, and professional development experiences. An analysis of data for the THEC STEM Professional Development Programs indicated participants overall experienced significant growth in all of these areas. Findings for each of these constructs are presented in Tables ES2-ES10 below.

Table ES3. Teacher Survey Findings: Teacher Opinions

Construct	Baseline % Agreement	End % Agreement
Students generally learn science/math best in classes with students of similar abilities.	55%	50%
I feel supported by colleagues to try out new ideas in teaching science/math.	83%	88%
Science/math teachers in this school have a shared vision of effective science/math instruction.	68%	72%
Science/math teachers in this school regularly share ideas and materials related to science/math.	71%	77%
Science/math teachers in this school are well supplied with materials for investigative science/math instruction.	43%	50%
I have time during the regular school week to work with my peers on science/ math curriculum and instruction.	39%	44%
I have adequate access to computers for teaching science/math.	48%	51%
I enjoy teaching science/math.	93%	93%
The science/math program in this school is strongly supported by local organizations, institutions, and/or business.	23%	37%

Table ES4. Teacher Survey Findings: Teacher Perceived Importance

Construct	Baseline % Agreement	End % Agreement
Provide concrete experiences before abstract concepts.	79%	82%
Develop students' conceptual understanding of science/math.	90%	92%
Take students' prior understanding into account when planning curriculum and instruction.	88%	90%
Make connections between science/math and other disciplines.	83%	86%
Have students work in cooperative learning groups.	78%	80%
Have students participate in appropriate hands-on activities.	89%	91%
Engage students in inquiry-oriented activities.	82%	88%
Have students prepare project/ laboratory/research reports.	48%	57%
Use computers.	58%	67%
Engage students in application of science/math in a variety of contexts.	80%	82%
Use performance-based assessment.	67%	71%
Use portfolios.	35%	39%
Use informal questioning to assess student understanding.	81%	86%

Table ES5. Teacher Survey Findings: Instructional Influences – Encourages Effective Instruction

Construct	Baseline % Agreement	End % Agreement
State and/or district curriculum frameworks.	50%	54%
State and/or district testing policies and practices.	33%	33%
Quality of available instructional materials.	46%	55%
Access to computers for science/math instruction.	44%	53%
Funds for purchasing equipment and supplies for science/math.	29%	36%
System of managing instructional resources at the district/school level.	30%	41%
Time available for teachers to plan and prepare lessons.	36%	49%
Time available for teachers to work with other teachers.	35%	45%
Time available for teacher professional development.	45%	54%
Importance that the school places on science/math.	63%	64%
Consistence of science/math reform efforts with other school/district reforms.	41%	45%
Public attitudes toward reform.	22%	27%

Table ES6. Teacher Survey Findings: Teacher Preparedness

Construct	Baseline % Agreement	End % Agreement
Provide concrete experiences before abstract concepts.	71%	88%
Develop students' conceptual understanding of science/math.	74%	91%
Take prior understanding into account when planning curriculum & instruction.	81%	92%
Make connections between science/math and other disciplines.	72%	89%
Use of cooperative learning groups.	80%	91%
Have students participate in appropriate hands-on activities.	77%	94%
Engage students in inquiry-oriented activities.	56%	86%
Have students prepare project/ laboratory/research reports.	35%	66%
Use computers.	66%	85%
Engage students in applications of science/math in a variety of contexts.	60%	89%
Use performance-based assessment.	69%	85%
Use portfolios.	32%	50%
Use informal questioning to assess student understanding.	82%	91%
Lead a class of students using investigative strategies.	58%	85%
Manage a class of students engaged in hands-on/project-based work.	75%	93%
Help students take responsibility for their own learning.	72%	89%
Recognize and respond to diversity.	75%	87%
Encourage students' interest in sci/math.	82%	94%
Use strategies that specifically encourage participation of females/minorities.	56%	80%

Table ES7. Teacher Survey Findings: Frequency of Use of Effective Pedagogy

Construct	Baseline % Agreement	End % Agreement
Introduce content through formal presentations.	72%	71%
Arrange seating to facilitate student discussion.	75%	83%
Use open-ended questions.	81%	88%
Require students to supply evidence to support their claims.	70%	85%
Encourage students to explain concepts to one another.	77%	83%
Encourage students to consider alternative explanations.	68%	79%
Allow students to work at their own pace.	70%	75%
Help students see connections between science/math and other disciplines.	66%	77%
Use assessment to find out what students know before or during a unit.	63%	68%
Embed assessment in regular class activities.	81%	83%
Assign science/math homework.	62%	60%
Read and comment on the reflections students have written in their notebooks or journals.	28%	43%

Table ES8. Teacher Survey Findings: Student Activities

Construct	Baseline % Agreement	End % Agreement
Participate in student-led discussions.	48%	65%
Participate in discussions with the teacher to further science/math understanding.	78%	85%
Work in cooperative learning groups.	81%	84%
Make formal presentations to the class.	17%	28%
Read from a science/math textbook in class.	36%	36%
Read other science/math-related materials in class.	56%	46%
Review homework/worksheet assignments.	64%	65%
Work on solving a real-world problem.	68%	70%
Share ideas or solve problems with each other in small groups.	66%	76%
Follow specific instructions in an activity or investigation.	62%	71%
Design or implement their own investigation.	17%	32%
Work on models or simulations.	20%	35%
Work on extended science/math investigations or projects.	10%	23%
Participate in field work.	5%	13%
Record, represent, and/or analyze data.	28%	42%
Write reflections in a notebook/journal.	33%	48%
Work on portfolios.	10%	16%
Take short-answer tests.	50%	47%
Take tests requiring open-ended responses.	37%	45%

Table ES9. Teacher Survey Findings: Parental Support

Construct	Baseline % Agreement	End % Agreement
Volunteer to assist with class activities.	2%	6%
Donate money or materials for classroom instruction.	8%	11%
Attend parent-teacher conferences.	33%	37%
Attend school activities such as PTA meetings and Family Science/Math nights.	9%	13%
Voice support for the use of an investigative approach to science/math.	5%	9%
Voice support for traditional approaches to science/math instruction.	9%	12%

Table ES10. Teacher Survey Findings: Principal Support

Construct	Baseline % Agreement	End % Agreement
Encourages selection of science/math content and instructional strategies to address individual students' learning.	81%	81%
Accepts the noise that comes with an active classroom.	84%	85%
Encourages the implementation of current national standards in science/ math education.	84%	86%
Encourages innovative instructional practices.	88%	89%
Enhances the science/math program by providing me with needed materials and equipment.	56%	60%
Provides time for teachers to meet and share ideas with one another.	58%	64%
Encourages me to observe exemplary science/math teachers.	44%	50%
Encourages me to make connections across disciplines.	74%	79%
Acts as a buffer between teachers and external pressures.	70%	68%

Table ES11. Teacher Survey Findings: Professional Development Experiences

Construct	Baseline % Agreement	End % Agreement
Participating in PD has increased my science/math content knowledge.	21%	39%
Participating in PD has increased my understanding of how children think about and learn science/math.	25%	39%
Participating in PD has increased my ability to implement high-quality science/math instructional materials.	26%	40%

Program-Developed Pre/Post Content Assessments

The analysis of data provided by THEC STEM Professional Development programs revealed significant growth in STEM content knowledge.

Individual Program-level Findings

In addition to the overall THEC STEM PD collective program analysis, individual program analyses were conducted and narratives for each funded program have been included in previous reports. Sixteen of the 29 funded programs realized significant growth in all aspects of teacher quality and content knowledge.

The programs that have been determined to represent best practice in STEM PD for the state of Tennessee include the following, presented alphabetically by round.

Round One Programs

1. Austin Peay University (APSU) – Grades 3-5 Mathematics (Principal Investigators Assad and Wells)
2. East Tennessee State University (ETSU) – Grades 3-5 Science (Principal Investigators Tai and Ho)
3. East Tennessee State University (ETSU) – High School Chemistry (Principal Investigators Rhoton and Zhao)
4. Tennessee Technological University (TTU) – Grades 3-5 Science (Principal Investigators Gore and Hunter)
5. Tennessee Technological University (TTU) – High School Chemistry (Principal Investigators Rust and Stevens)

Round Two Programs

1. East Tennessee State University (ETSU) – High School Chemistry & Physics (Principal Investigators Rhoton and Zhao)
2. Lipscomb University (LU) – Grades 4-7 Mathematics and Science (Principal Investigators Wells, Morel & Nelson)
3. Lipscomb University (LU) – High School Algebra (Principal Investigators Nelson and Thornthwaite)
4. Middle Tennessee State University (MTSU) – Grades 4-8 Mathematics and Science (Principal Investigators Kimmins and Winters)
5. Middle Tennessee State University (MTSU) – High School Mathematics and Science (Principal Investigators Strayer and Brown)
6. Tennessee Technological University (TTU) – Grades 3-6 Mathematics and Science (Principal Investigators Pardue and Howard)
7. Tennessee Technological University (TTU) – High School Mathematics and Science (Principal Investigators Fidan and Baker)
8. Tennessee Technological University (TTU) – K-2 Mathematics and Science (Principal Investigators Baker and Fromke)
9. Tennessee Technological University (TTU) and Roane State Community College (RSCC) – Middle School Mathematics and Science (Principal Investigators Suters and Lee)
10. University of Tennessee at Chattanooga (UTC) – Grades 6-12 Science (Principal Investigators Ingraham, Ellis, and Carver)
11. University of Tennessee at Martin (UTM) – Grades 5-9 Science (Principal Investigators Cox and Withmer)

SUMMARY

Overall, the evaluation of the THEC STEM PD programs revealed significant growth in science and mathematics teacher effectiveness and attitudes. At an individual program level, 16 of the 29 total programs realized significant growth in teacher quality and content knowledge measures and should be considered as best practice models for the state of Tennessee.

I. INTRODUCTION

BACKGROUND

In April 2011 the Tennessee Higher Education Commission (THEC) released a request for proposals (RFP) for the first round of Race to the Top funded STEM Professional Development (PD) programs. Eleven programs were funded across the state of Tennessee in Round One. In spring, 2012, the second call for proposals was released, and 18 additional programs were funded. In total, 29 programs were funded, with the intent of promoting innovative practices in STEM (science, technology, engineering, and mathematics) education by further developing K-12 STEM teachers' pedagogical skills and content knowledge. (See Table 1.) In addition, the PD programs funded through this grant program and determined to be highly effective may be shared throughout the Tennessee STEM Innovation Network (TSIN). Highly effective programs are defined as those that have significant gains in teacher pedagogical skills and content knowledge.

The primary objectives of the program are:

1. To deliver high quality, research-based STEM professional development to K-12 teachers to improve pedagogical skills and content knowledge.
2. To align with the goals of Tennessee's First to the Top plan, including School readiness, College and Career readiness, Implementing the Common Core Standards, and Postsecondary Access and Success.
3. To create a STEM Professional Development best-practices warehouse for use throughout Tennessee's STEM Innovation Network (TSIN) to ensure sustainability of this PD beyond funding from Race to the Top. Through replication and sustainability, it is intended that those PD programs that are models of good practice will and can be accessed and replicated widely throughout the TSIN in order to foster deeper learning of STEM content knowledge for all students.

This final evaluation report will focus on the complete analysis of data collected for both Round One and Round Two of the THEC STEM Professional Development program.

Table 1. THEC STEM Professional Development Programs Round One and Round Two

Round One THEC STEM Professional Development Programs

Institution and Program Title	Number of Teachers*	Focus Area
APSU, *Momentum*	30	ES mathematics
ETSU, *MICH: Modeling Instruction in High Schools*	20	HS science
ETSU, *Reaching for Excellence in Elementary School Science*	20	ES science
LU, *Hands-on Chemistry*	17	HS science
MTSU, *Project EMPOWER*	43	HS mathematics
TTU, *Developing Middle School Math Teachers' PCK*	30	MS mathematics
TTU, *Embedding Inquiry & Technology*	24	ES science
TTU, *Numeracy and Multiple Representations for Grades 1-3 Teachers*	29	ES mathematics
TTU, *Transforming Matter and Classrooms*	10	HS science
UTC, *Numeracy, Representation, and STEM Connections for K-2*	29	ES Math
UTC, *TELMU*	29	MS mathematics

Round Two THEC STEM Professional Development Programs

Institution and Program Title	Number of Teachers*	Focus Area
ETSU, I*ncorporating Active Learning into Life Sciences Teaching*	18	MS mathematics and science
ETSU, *Integrating Hands-on STEM Activities with Math and Reading CCSS*	24	EC mathematics and science
ETSU, *PCMI*	25	HS science
ETSU, *Project SEE*	25	EC science
LU, *Functions of Algebra*	20	HS mathematics
LU, *Integrating STEM: The Power of Science*	10	HS science
LU, *Making Mathematics Matter*	20	MS mathematics and Science
MTSU, *StaRT*	35	HS mathematics and science
MTSU, *UC STEM*	25	MS mathematics and science
TTU, *Designing the Future*	25	HS mathematics and science
TTU, *Shaping Early STEM Learning*	29	EC mathematics and science
TTU, *STEM Around Us*	35	ES mathematics and science
TTU and RSCC, *From Earth to Space with STEM*	30	MS mathematics and science
UM, *mMind*	29	MS/HS mathematics and science
UM, *Professional Development for Grades 5-8*	28	MS mathematics and science
UM, *Water, Water Everywhere*	18	MS science
UTC, *Learning Science through Writing*	23	MS/HS science
UTM, *STEM Integration for Middle School Teachers Academy*	28	MS science

* Numbers presented in Table 1 reflect the number of teachers who actually completed each program. This number does not equate, however, to the number of individuals who participated in data collection activities as those numbers vary by activity.

II. RESEARCH METHODS

RESEARCH QUESTIONS

Three research questions, listed below, guided this evaluation. All are aligned with the primary objectives of the THEC STEM PD Program:

1. What impact, if any, do THEC STEM professional development programs have on teachers' pedagogical skills and STEM content knowledge?
2. What impact, if any, do THEC STEM professional development programs have on teachers' opinions regarding the teaching of STEM?
3. Which funded STEM professional development programs demonstrate significant growth in Teacher Quality (pedagogical skills and content) and should be considered for inclusion as best practice for Tennessee?

CORE CONCEPTUAL FRAMEWORK

Much has been learned through recent attempts at designing professional development programs for STEM teachers. As the knowledge base on educational reform and improving teacher quality has grown over the past decade (e.g., Johnson, Kahle, & Fargo, 2007a, 2007b; Johnson & Fargo, 2010; Loucks-Horsley, Love, Stiles, Mundry, & Hewson, 2007; Putnam & Borko, 1997) it has become more evident that traditional professional development formats do not result in sustained change in practice. Professional development linked to state and/or district reform initiatives have demonstrated the ability to transform educational practice systemically (Desimone, 2009). However, since enactment of the *No Child Left Behind Act of 2001* few attempts have been made to explore the ability of effective teacher quality programs to achieve systemic reform (Desimone, 2009; Johnson et al., 2007b).

Desimone (2009) published a seminal paper wherein she conducted a rigorous review of empirical studies of professional development to produce a core conceptual framework for research-based, effective professional development, defined as models that have had positive impact on "increasing teacher knowledge and skills and improving their practice, which hold promise for student achievement" (p. 183). The components of the core conceptual framework include content knowledge focus, active learning experiences, coherence with state/district goals and standards, extended duration of a program across academic year(s), and collective participation of teams of teachers from same school.

THEC required all submitted proposals to include these five core components in the design of their programs. All funded PD projects included the core components as the basis into which they inserted their content and context.

In most of the published research on professional development in small settings, it has taken at least two years before significant change in teacher effectiveness has been realized. The THEC STEM PD program has provided the setting for the first large-scale implementation of the research-based core conceptual framework for effective professional development. Moreover, Tennessee has taken steps to integrate research into the significant Race to the Top investment, and the evaluation of the THEC STEM PD program will provide much-needed insight into educational reform.

DATA COLLECTION AND ANALYSIS

The evaluation of the THEC STEM PD programs included a variety of qualitative and quantitative data to investigate the impact of THEC STEM PD. The data collection and analysis activities for this report included teacher classroom observations in digital recording format and two surveys completed by participating teachers. Each of these is described in more detail below.

Teacher Observation Data

Teacher observations were conducted for use in determining potential increased use of STEM pedagogical skills and STEM content knowledge for THEC STEM PD participants. Each participating teacher in all funded STEM PD programs was asked to submit three digital recordings of an appropriate STEM lesson. The first recording was to be conducted prior to beginning participation in the THEC STEM PD program. The second was to occur at the mid-point of participation (August 2012) and the final recording was to be completed and submitted by December 2012.

Classroom Observation Instrument

The Local Systemic Change (LSC) Classroom Observation Protocol is an observation tool used to assess the degree of instructional reform in math and science. The LSC protocol was developed by Horizon Research for use with the National Science Foundation's (NSF) funded State Systemic Initiatives (SSI) as a measure of reform-based instructional practices. The LSC Classroom Observation Protocol is being used as the measure of growth in teacher pedagogical skill use and is one measure of teacher content knowledge for the THEC STEM PD program. The LSC tool is valid for use in this evaluation based on the research-based foundation and wide-scale implementation of the LSC protocol in many empirical studies. Using the LSC, teacher instruction is observed and given ratings on 32 items included in four domains (see Table 2).

Table 2. LSC Domains

Domain	Number of Items
Design of Lesson	10
Implementation of Lesson	7
Classroom Culture	6
Math/Science Content	9

The *Design of Lesson* domain focuses on the structure of the observed lesson and investigates a variety lesson considerations such as the sequencing of instructional activities, roles of students and teachers, resources available, eliciting of prior knowledge, time provided for sense making, attention to diversity, and collaborative learning. The *Implementation of Lesson* domain examines the use of investigative STEM strategies employed by the teacher, as well as the pace of the lesson, attention to student understanding, questioning strategies, and both formative and summative assessments. The *Classroom Culture* domain assesses a teacher's ability to create and facilitate a classroom environment, which supports active participation, respect for ideas, effective collaboration, and inquiry into student ideas, questions, and real-world connections. The *Mathematics/Science Content* domain examines teacher understanding of content, as well as appropriateness of the level of content included in the lesson, the level of student engagement with content, and interdisciplinary and real-world connections presented by the teacher.

Each item within each domain ranges is scored on a scale of 0 to 5, with 0 being used when there is no evidence of a component within a domain, and a score of 5 awarded when a component is used "to a great extent". Each domain has multiple questions that are scored individually, and an overall rating (i.e., mean score) for each domain is generated (see Table 3).

Table 3. LSC Overall Rating

Score	Title
0-1.9	Ineffective Instruction
2-2.9	Elements of Effective Instruction
3-3.9	Beginning of Effective Instruction
4-4.9	Accomplished, Effective Instruction
5	Exemplary Instruction

An overall score of 0 to 1.9 is characterized with a rating of *Ineffective Instruction*. The LSC protocol describes this as a classroom where there is "little or no evidence of student thinking or engagement with important ideas of mathematics/science. Instruction is highly unlikely to enhance students' understanding of the discipline or to develop their capacity to successfully do mathematics or science". With this rating, the delivered lesson is characterized as either *passive learning* or *activity for* *activity's sake.* Passive learning is when students are passive recipients of information from the teacher or textbook. Activity for activity's sake happens when a hands-on activity is employed with no clear purpose and does not lead to student conceptual development of STEM.

An overall score of 2-2.9 receives the rating of *Elements of Effective Instruction.* The LSC protocol describes this as a classroom where "instruction contains some elements of effective practice but there are serious problems in the design, implementation, content, and/or appropriateness for many students in the class". Examples of this are inappropriate content and/or level of content, lack of ability to address student difficulties, lack of opportunities for inquiry and investigation of student ideas, and problem solving.

An overall score of 3-3.9 is classified as *Beginning Stages of Effective Instruction.* The LSC protocol describes this as a classroom where, "instruction is purposeful and characterized by quite a few elements of effective practice". In this classroom, students are engaged in meaningful work at times but there are still a few weaknesses with the delivery of the lesson.

An overall score of 4-4.9 is characterized as *Accomplished, Effective Instruction.* The LSC protocol describes this as a classroom that is, "purposeful and engaging for most students". Students are engaged in meaningful work, including investigations, and the lesson is well designed and implemented. Some limitations in ability to adapt content and/or pedagogy still exist and ability to respond to student needs is also limited. Instruction is "*quite likely"* to enhance student ability to do STEM.

An overall score of 5 is *Exemplary Instruction.* The LSC protocol describes this a classroom where, "*purposeful instruction* [is occurring] *and all students are highly engaged most or all of the time with meaningful work".* The lesson is "*artfully implemented";* the teacher is flexible and responds to student needs and interests; and instruction is highly likely to enhance student understandings of the discipline and to develop their capacity to do STEM.

Response Rate - Teacher Observation Data

Collectively, for Round One and Two, 667 teachers were observed at least once. Of those 667 teachers, 236 teachers (35.4 percent) completed two full observations, which were then scored, and 270 teachers (40.5 percent) completed and had scored three full observations. These teachers will serve as the sample for this report, as they participated in the entire PD program and provide the most accurate measure of change over time. Participants from the two rounds were unevenly distributed, however, with 67.4 percent (n=182) coming from Round Two, and 32.6 percent (n = 88) from Round One.

Analysis of Teacher Observation Data

Teacher videos were rated by a team of evaluators and analyzed quantitatively. All videos were viewed and scored by two independent raters using the LSC Classroom Observation Protocol in four domains, including design of lesson, implementation of lesson, mathematics/science content knowledge, and classroom culture, as well as an overall rating. This measure is used to determine improvement in teacher pedagogical skills and content knowledge as demonstrated through actual teacher practice.

Total scores for each domain were computed. Each domain section was comprised of a different number of total items (see Table 2). Individual item ratings ranged from 1-5 with 1 being lowest and 5 being highest (see Table 3). In addition to the domain rating, an overall rating was also assigned to each teacher for each lesson. To assess teacher growth in specific classroom practices over time and by program classification (high school chemistry, elementary science, primary math, elementary math, middle grades math) a 3-Within, 7-Between Repeated Measures ANOVAs with post-hoc investigation for differences at each observation time and between program classification groups was conducted. Finally, growth examinations between all observation time points using 3-Within Repeated Measures ANOVAs with post-hoc investigation for each specific program's STEM Teacher Quality results are conducted. Since sample sizes for individual programs are small, one-tailed tests were run to increase the sensitivity for finding statistically significant differences over time.

Teacher Survey Data

Two measures were used in this evaluation to determine teacher-reported growth in use of effective pedagogical skills, as well as potential change in opinions for participants in the funded THEC STEM PD programs. This data was in addition to classroom observation data, which also examined use of effective pedagogical content knowledge. Participants completed appropriate questionnaires for their grade band and content area. Participants also completed the surveys in a pre/post manner for the program online through Survey Monkey, prior to participation in the PD and at the end of the PD program.

Teacher Survey Instruments

Two surveys were used in this evaluation. The LSC Teacher Questionnaires (e.g., mathematics and science versions for K-8 and 9-12) were selected based upon their alignment with the LSC Classroom Observation protocol (used for the classroom observational data) and previous use in the NSF funded SSIs (http://www.horizon-research.com/LSC/news/heck_rosenberg_crawford_2006a.php). Additionally, the Survey of Enacted Curriculum (SEC), developed by the SEC Collaborative (https://secure.wceruw.org/seconline/secWebHome.htm), which has been used extensively in Georgia, Kansas, Kentucky, Michigan, Mississippi, and Ohio, is a second research-based instrument used for the evaluation. Collectively, the two instruments were used to measure preparedness to teach STEM, influences on instruction, beliefs regarding STEM teaching, parental and principal support, and quality of PD experiences.

Response Rate - Teacher Survey

A total of 452 teachers from the 29 THEC STEM PD programs completed both a pre- and post-survey. These 452 teachers serve as the sample for this report. Of this sample, 146 participants (32.3 percent) were from Math K-8 programs, 148 participants (32.7 percent) were from Science K-8 programs, 45 participants (10.0 percent) were from Science 9-12 programs, and 113 teachers (25.0 percent) were from Math 9-12 programs.

Analysis of Teacher Survey Data

A 2-between 2-within Factorial ANOVA was employed to assess overall growth from pre/post regardless of the PD group and also look for differences in growth by PD content area (science vs. math). Next, multiple Chi-Square Tests of Independence were employed to examine pre- to post-survey response percent growth for individual items regardless of the PD program. Finally, because it is very difficult to change teacher beliefs and perceptions, one-tailed tests were implemented to increase the power for finding statistical differences. Further, we considered any pre/post improvement at the $p < .10$ to be statistically significant.

Teacher Content Assessments

Each program was asked to develop their own content assessments (25 items as requested by the RFP) to determine participant growth in content knowledge. Each program submitted copies of assessments, keys, and a spreadsheet with individual teacher responses to each item for pre/post. Some programs did not follow the guidelines for assessments and data from those programs were not in a format that would fit analysis for this evaluation and were not included.

Content Assessment Instrument

Each professional development program created their own assessment of teacher content knowledge aligned with content and grade levels covered in their individual program. As a result, all teacher content knowledge assessment items are different across tests. However, all assessment developers were to follow the same guidelines when creating and distributing tests: 1) pre- and post-test items given to teachers should consist of the same items on both tests; 2) all items should be objective type items (scored as correct/incorrect rather than subjectively scored with a rubric); 3) assessments should be comprised of 25 items; and 4) teachers needed the same identification number in each pre- and post-test files to allow for pre/post content knowledge comparison. Most of the eleven round one and two programs followed these guidelines with the exception of three programs which used subjectively scored items (programs 7 and 9), a differing number of pre- and post-test items (program 3), did not identify teachers with the same code in pre- and post-test files (program 3), or did not submit data for analysis (program 16). As such, data from programs 3, 7, 9, and 16 were not included in analyses because they did not follow the assessment creation and distribution guidelines in ways that made comparison of pre/post teacher content knowledge results impossible. While some programs distributed more or less than 25 items on their assessments, participants in these groups were not eliminated from analysis because percentage correct was used as the metric for comparison rather than total number of items correct.

Regardless of which Tennessee Race to the Top STEM PD program teachers were involved in, teachers' math/science content knowledge significantly improved from pre-test (M = 59.33%, SD = 20.71%) to post-test (M = 73.29%, SD = 17.85%); $t(521) = 19.57$, $p < .000$. The effect size is considered large ($_{p}^{2}$=.424) with 42.4 percent of the variance in teacher content knowledge accounted for by time of the test. The overall teacher pre- and post-test average content knowledge percent correct growth over the program was from 59.33 percent correct at baseline to 73.29 percent correct at end of program. A one-way ANOVA was used to analyze program developed content knowledge assessment data by type of program.

Limitations

All quantitative research is subject to limitations from methodological threats to internal and external validity (Onwuegbuzie, 2000). Internal validity focuses on the research design and asks if it is appropriate to support the differences found in the dependent variable as a result of the independent variable and nothing else. External validity addresses a study's ability to generalize findings from one study to and across populations, settings, and times. For this evaluation study, two major methodological limitations to validity are acknowledged: 1) teacher participation in data collection, and 2) nature of the content knowledge tests.

Teacher participation in data collection is a potential external validity limitation in this evaluation study. Out of 733 total participating teachers in both rounds of the THEC STEM PD programs, response rates for completing the teacher survey at least once was 81.4 percent (n=250), having one classroom observation performed was 82.1 percent (n=252), and 72.3 percent (n=222) completed the program developed content knowledge assessment for teachers (pre/post). While these overall response rates are high, when considering that this evaluation was of a longitudinal nature, the response rates are not quite as impressive. Only 54.1 percent (n=166) of participating THEC STEM PD teachers completed both pre- and post-surveys, 40.4 percent (n=124) had three full classroom observations recorded, and 52.4 percent (n=161) produced usable pre/post achievement test scores. Further, because some THEC STEM PD participants did not participate in the data collection process, findings of this evaluation are vulnerable to non-response error. Non-response error may occur when a significant number of THEC STEM PD teachers choose to not respond and these non-respondents are significantly different from those THEC participants who responded and thus the results may become non-generalizable to the larger THEC STEM PD program sample. Any time a response rate is under 60-70 percent non-response needs to be examined further. In this evaluation, THEC participant demographics (e.g., program content, program grade level focus, gender, as ethnicity) for those responding to data collection procedures are similar to that of the overall THEC participant group. As such, we can say that there does not appear to be any systematic non-response issues making this a lesser concern than if there were specific sub-groups of individuals choosing to not participate.

The nature of the program developed content knowledge tests for teachers is an internal limitation for this evaluation study. All content knowledge tests were developed by the individual professional development programs to focus on the specific content each program was covering. While this does allow for greater content validity for these assessment outcomes, there is limited (if any) comparability across assessments. Thus, there is no way of knowing if one assessment was significantly more challenging or easier than another assessment. Consequently, comparability of growth from pre/post across programs attributing differences to type of PD delivered is certainly confounded by the differences in tests and should be done with extreme caution. It is acceptable to look at growth from pre/post for an individual program, but comparing one program's growth to another may have little to do with the PD implemented and more to do with the assessment used.

III. FINDINGS OVERALL FOR THEC STEM PD INVESTMENT – RESEARCH QUESTION 1

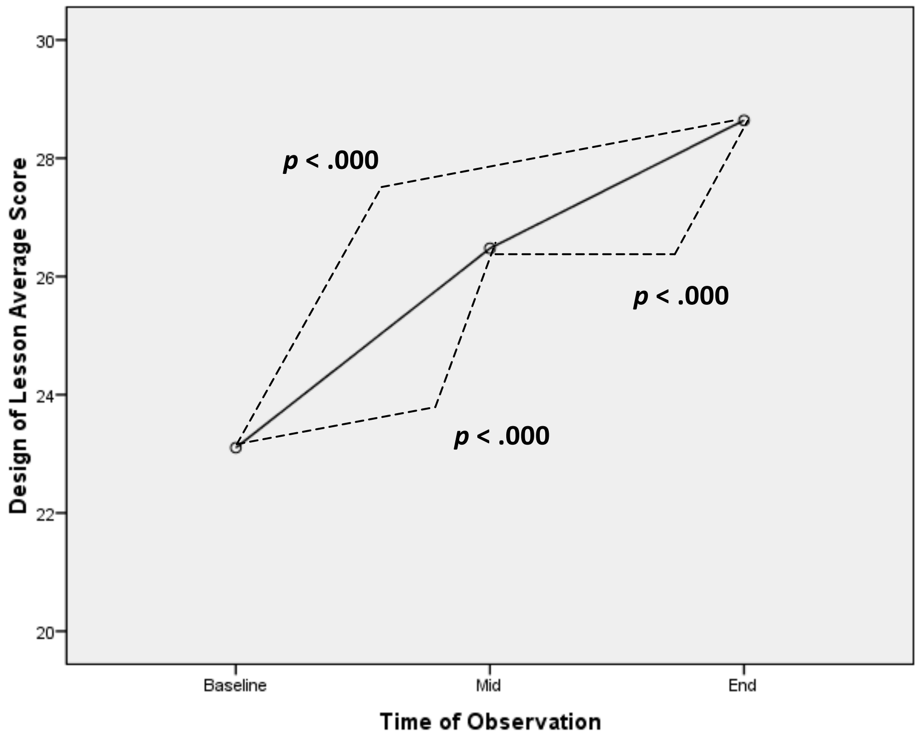
CLASSROOM OBSERVATION FINDINGS

The Local Systemic Change Classroom Observation Protocol (LSC) was used to examine teacher observations in four key areas: design of lesson, implementation of lesson, culture of instruction, and content knowledge delivered. Analysis of these videos revealed significant improvement in all four areas as indicated by findings presented below.

Design Of Lesson

An analysis of data for the THEC STEM PD programs (29 total programs) indicated there was significant growth in the Design of Lesson construct, which encompasses the extent of planning, organization, resources, equity, collaboration, flow, assessments, and sense making that takes place in the lesson delivery. At baseline, the mean score average (2.39) was rated a Level 2: Elements of Effective Instruction (*M* = 23.87, *SD* = 5.24), which increased to 2.67 (*M* = 26.67, *SD* = 4.84) at the second observation point midway through the professional development program, and increased further to (average score of 2.88) at the final observation (*M* = 28.82, *SD* = 5.35), $F(2) = 39.59$, $p < .000$. The effect size is considered large ($_p^2=.132$), with 13.2% of the variance in Design of Lesson scores accounted for by time of the observation. Figure 1 shows the statistically significant overall increase in average Design of Lesson scores over time.

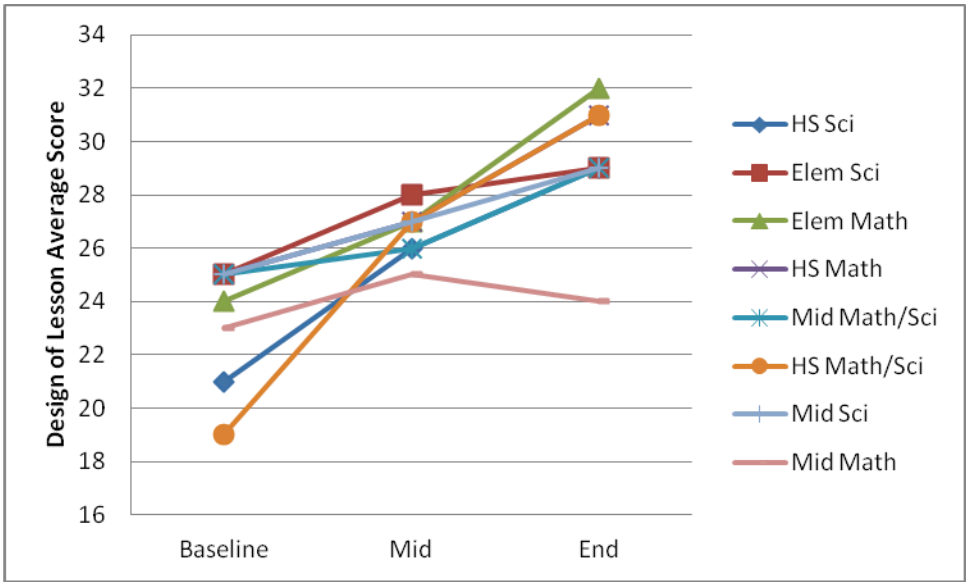
Figure 1. Design of Lesson Average Score Over Time for THEC STEM PD Programs



Average scores could have an overall range of 10-50, since there are 10 items on a 5-point scale in this sub-section. Statistically significant increases were noted between all observation points.

State level findings did vary by type of program (e.g., mathematics, science, or grade range), meaning there was a statistically significant difference in design of lesson between program classifications, $F(6) = 4.31$, $p < .000$. The effect size is considered medium ($\eta_p^2 = .117$), with 11.7 percent of the variance in design of lesson score accounted for by type of program. The only significant differences in program type were between Middle School Math programs, which were significantly lower compared to Elementary Science ($p < .000$), Elementary Math ($p < .05$), High School Math ($p < .01$), and Middle School Math/Science ($p < .05$). The average design of lesson score across time ranged from 1.85 (High School Math/Science) to 3.20 (Elementary Math), which are equivalent to a Level 1: Ineffective Instruction and Level 3: Beginning stages of Effective Instruction respectively. There was also a statistically significant interaction between program classification and time of observation for design of lesson, $F(12) = 2.50$, $p < .001$. This means as time went on, the group overall improved. The effect size is considered medium ($\eta_p^2 = .071$), with 7.1 percent of the variance in design of lesson score accounted for by the interaction between observation time and program classification. Figure 2 shows that all program classifications increased in design score from baseline to mid-program observations and again increased from mid- to end-of-program observations.

Figure 2. Design of Lesson Average Score Over Time by Program Classification for THEC STEM PD Programs

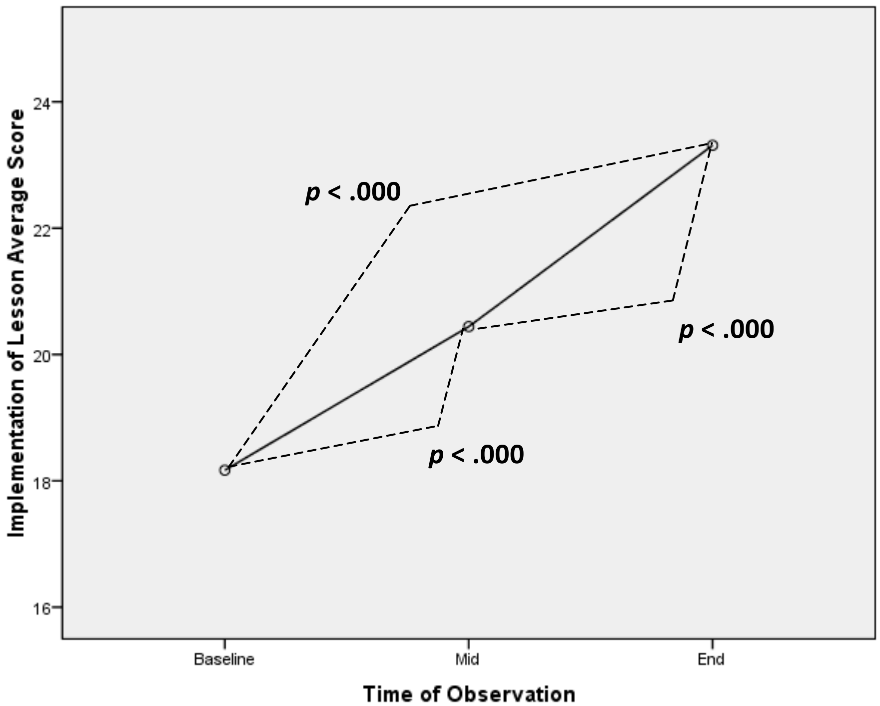


Average scores could have an overall range of 10 -50, since there are 10 items on a 5-point scale in this sub-section. The only significant differences noted over time were between Middle School Math programs, which were significantly lower compared to Elementary Science ($p < .000$), Elementary Mathematics ($p < .05$), High School Math ($p < .01$), Middle School Mathematics, and Middle School Mathematics/Science ($p < .05$).

Implementation Of Lesson

Regardless of program classification, teachers involved in Tennessee's Race to the Top STEM PD schools significantly improved their Implementation of Lesson scores from their average baseline rating of 2.61 or a Level 2: Elements of Effective Instruction ($M = 18.30$, $SD = 4.12$), to a average rating of 2.92 ($M = 20.41$, $SD = 4.26$) at the second observation recorded at the mid-point of the professional development program, scores finally rose to an average rating of 3.30 or a Level 3 at the end-point observation ($M = 23.11$, $SD = 4.51$), $F(2) = 37.91$, $p < .000$. The implementation of lesson construct considers the level of investigative mathematics/science in the lesson, quality of classroom management strategies, pace of the lesson, ability to modify instruction based upon student understanding, teacher questioning strategies, and formative assessments. The effect size is considered large ($_p^2 = .226$) with 22.6 percent of the variance in Implementation of Lesson scores accounted for by time of the observation. Figure 3 shows the statistically significant increase in average Implementation of Lesson scores over time.

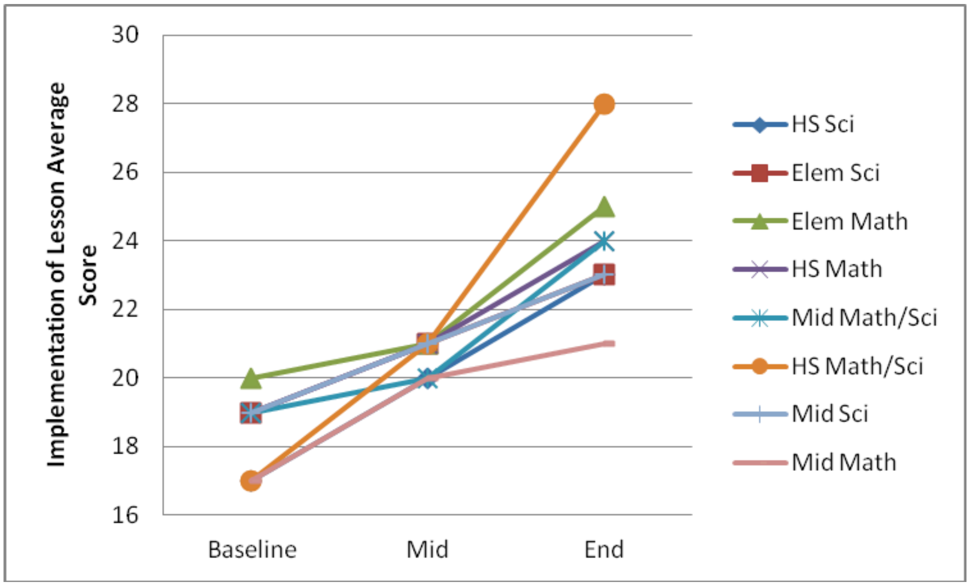
Figure 3. Implementation of Lesson Average Score Over Time for THEC STEM PD Programs



Average scores could have an overall range of 7-35, since there are seven items on a 5-point scale for this sub-section. Statistically significant increases were noted between all observation points.

State level findings did not vary by type of program (e.g., mathematics, science, or grade range) in terms of Implementation of Lesson, meaning there was not a statistically significant difference in implementation of lesson between program classifications, $F(6) = 1.94$, $p > .05$. The effect size is considered small ($\eta_p^2 = .057$), with 5.7 percent of the variance in implementation of lesson score accounted for by type of program. The average implementation of lesson score across time ranged from 2.39 (High School Chemistry) to 4.00 (High School Math/Science), which are equivalent to a Level 2: Elements of Effective Instruction and Level 4: Accomplished, Effective Instruction respectively. There was also not a statistically significant interaction between program classification and time of observation for design of lesson, $F(12) = 1.17$, $p > .05$. The effect size is again considered small ($\eta_p^2 = .035$), with 3.5 percent of the variance in implementation of lesson score accounted for by the interaction between observation time and program classification. Figure 4 shows that all program classifications increased in implementation score from baseline to mid-program observations and again increased from mid- to end-of-program observations, and none of the increases were significantly different by program over time.

Figure 4. Implementation of Lesson Average Score Over Time by Program Classification for THEC STEM PD Programs

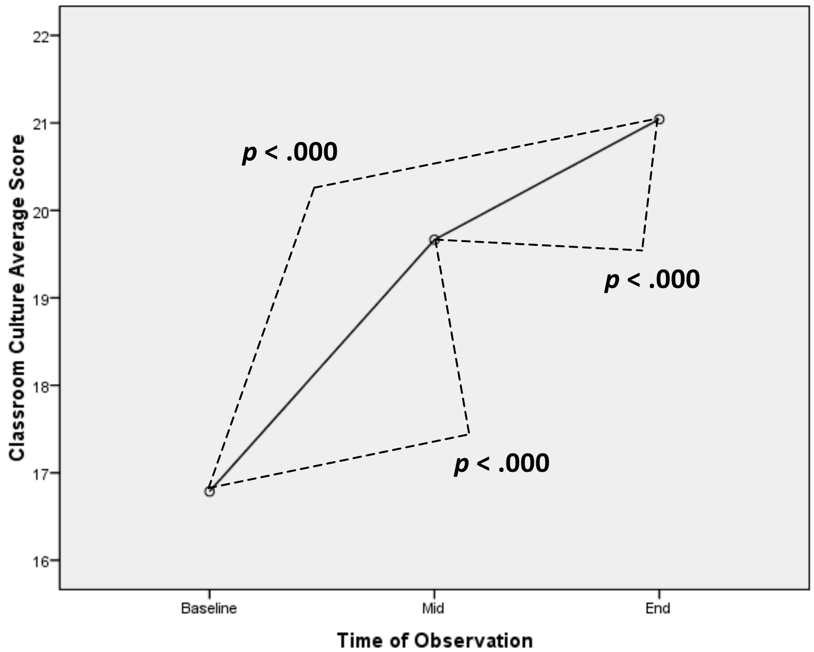


Average scores could have an overall range of 7-35, since there are seven items on a 5-point scale for this sub-section. No significant differences were revealed for programs across time by group.

Classroom Culture

The THEC STEM PD participants also significantly improved their overall Classroom Culture scores from baseline average rating of 2.84 or a Level 2: Elements of Effective Instruction (M = 17.03, SD = 4.06), to an average rating of 3.20 (M = 19.19, SD =3.89) on the second observation recorded at the mid-point of the professional development program. This rating increased to an average rating of 3.48 or a Level 3: Beginning Stages of Effective Instruction (M = 20.90, SD = 4.10) at the end-point observation, $F(2) = 51.85$, $p < .000$. The effect size is considered large (η_p^2 = .149), with 14.9 percent of the variance in Classroom Culture scores accounted for by time of the observation. Figure 5 shows the statistically significant increase in average Classroom Culture scores over time. Classroom Culture refers to the amount of active participation of all students and level of collaborative learning, including allowing students to explore their own ideas, questions, conjectures, and propositions or to challenge the ideas of others.

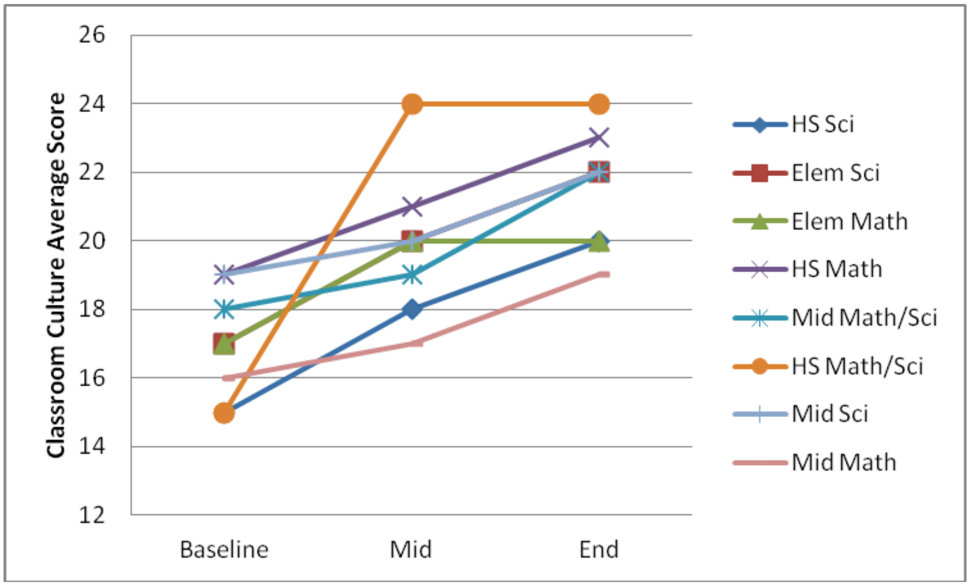
Figure 5. Classroom Culture Average Score Over Time for THEC STEM PD Programs



Average scores could have an overall range of 6-30 since there are six items on a 5-point scale for this sub-section. Statistically significant increases were noted between all observation points.

There was a statistically significant difference in classroom culture between program classifications, $F(6) = 6.52$, $p < .000$. The effect size is considered large ($\eta_p^2 = .150$), with 15.0 percent of the variance in classroom culture score accounted for by type of program. The only significant difference in program type were between High School Chemistry programs, which were significantly lower compared to High School Math ($p < .01$) and Middle School Math/Science ($p < .05$) programs. The average classroom culture score across time ranged from 2.42 (High School Math/Science) to 4.00 (High School Math/Science), which are equivalent to a Level 2: Elements of Effective Instruction and Level 4: Accomplished, Effective Instruction respectively. A statistically significant interaction between program classification and time of observation existed for classroom culture, $F(12) = 2.29$, $p > .05$. The effect size is considered small ($_p^2 = .058$), with 5.8 percent of the variance in Classroom Culture scores accounted for by the interaction of time of the observation and program classification. Figure 6 shows that all program classifications increased in classroom culture score from baseline to mid-observations and again increased from mid- to end-of-program observations.

Figure 6. Classroom Culture Average Score Over Time by Program Classification for THEC STEM PD Programs

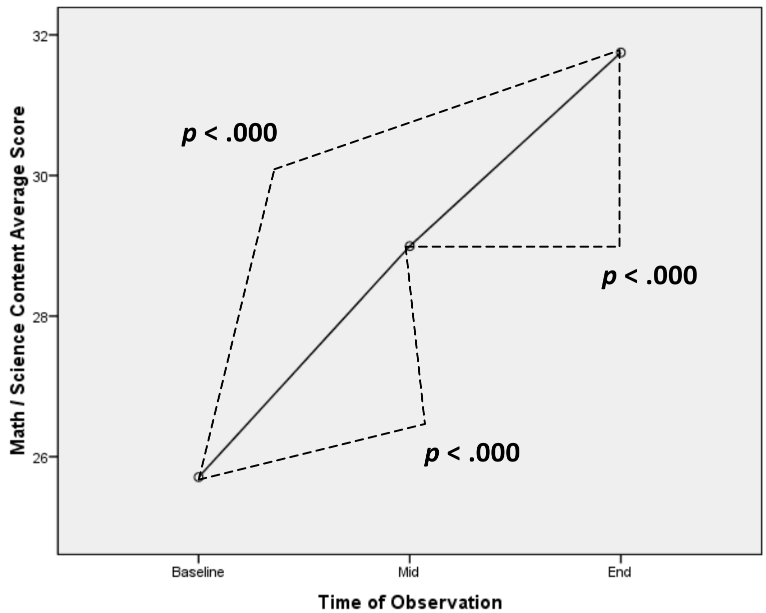


Average scores could have an overall range of 6-30 since there are six items on a 5-point scale for this sub-section. The only significant difference noted over time were between High School Chemistry programs, which were significantly lower compared to High School Math ($p < .01$) and Middle School Math/Science ($p < .05$) programs.

Mathematics/Science Content Domain

THEC STEM PD participants significantly improved their Mathematics/Science Content scores from a baseline score of 2.90, which is rated as a Level 2: Elements of Effective Instruction (M = 26.09, SD = 5.04), improving to an average rating of 3.22 (M = 29.00, SD = 4.99), which is rated as a Level 3: Beginning Stages of Effective Instruction at the second observation point mid-way through the professional development program. By the end of the program, participants experienced further growth, with an average score of 3.50 overall (M = 31.52, SD = 5.46), $F(2) = 56.25$, $p < .000$. The effect size is considered large ($_p^2$ = .159), with 15.9 percent of the variance in Mathematics/Science Content scores accounted for by time of the observation. Figure 7 shows the statistically significant increase in average Mathematics/Science Content scores over time.

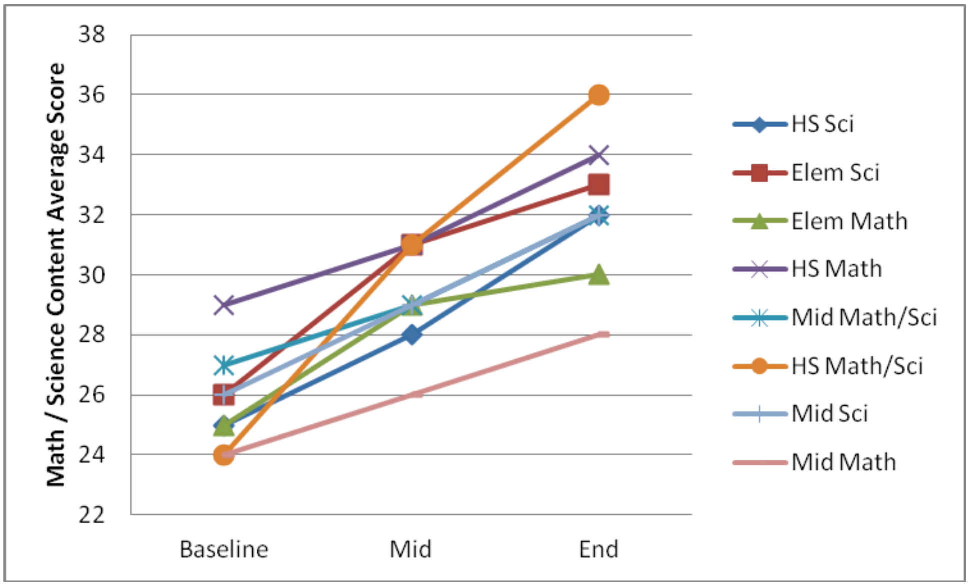
Figure 7. Mathematics/Science Content Domain Average Score Over Time for THEC STEM PD Programs



Average scores could have an overall range of 9 – 45 since there are 9 items on a 5-point scale for this sub-section. Statistically significant increases were noted between all time points.

There is a statistically significant difference in mathematics/science content between program classifications, $F(6) = 5.33$, $p < .000$. The effect size is considered medium ($\eta_p^2 = .126$), with 12.6 percent of the variance in classroom culture score accounted for by type of program. The only significant difference in program type was between High School Math programs, which were significantly higher compared to Elementary Math and High School Chemistry programs ($p < .05$). The average classroom culture score across time ranged from 2.61 (High School Math/Science) to 4.00 (High School Math/Science), which are equivalent to a Level 2: Elements of Effective Instruction and Level 4: Accomplished, Effective Instruction respectively. There is not a statistically significant interaction between program classification and time of observation for mathematics/science content, $F(12) = 1.35$, $p > .05$. The effect size is considered small ($_p^2 = .035$), with 3.5 percent of the variance in mathematics/science content scores accounted for by the interaction of time of the observation and program classification. Figure 8 shows that all program classifications increased in classroom culture score from baseline to mid-observations and again increased from mid- to end-of-program observations.

Figure 8. Mathematics/Science Content Average Domain Score Over Time by Program Classification for THEC STEM PD Programs



Average scores could have an overall range of 9-45, since there are nine items on a 5-point scale for this sub-section. The only significant difference in program type noted over time were between High School Math programs, which were significantly higher compared to Elementary Math and High School Chemistry programs ($p < .05$).

CONTENT KNOWLEDGE ASSESSMENT FINDINGS

For all types of PD programs there was statistically significant growth from pre- to post-test in terms of teacher content knowledge. While each program type showed statistically significant average increases from pre- to post-test in teacher content knowledge, Table 4 shows that teachers participating in High School Math/Science and Middle Grades Science programs appeared to make the greatest gains with each program type moving their teachers' content knowledge up approximately 30 percentage points from pre- to post-test (see Table 4). Further statistical analysis (One-Way ANOVA) revealed a statistically significant difference between groups in terms of pre-post teacher content knowledge growth; $F(7) = 13.71$, $p < .000$. Post-hoc analysis indicates that Elementary Science, High School Math/Science, and Middle School Science programs had significantly greater teacher content knowledge growth when compared to all other program types ($p < .01$).

Table 4. Pre- and Post-Average Percent Correct by Program Classification

Program Classification	Pre-test	Post-test	% Growth	Significant Growth
High School Science	73.86% (C)	80.61% (B)	+6.76% points	Yes (*p*<.000)
Elementary Science	58.08% (F)	78.06% (C)	+19.98 points	Yes (*p*<.000)
Elementary Mathematics	70.60% (C)	79.26% (C)	+8.66 points	Yes (*p*<.000)
High School Mathematics	52.04% (F)	62.88% (D)	+10.84 points	Yes (*p*<.000)
Middle Grades Mathematics/Science	52.43% (F)	66.01% (D)	+13.58 points	Yes (*p*<.000)
High School Mathematics/Science	47.27% (F)	79.32% (C)	+32.05 points	Yes (*p*<.000)
Middle Grades Mathematics	69.84% (D)	75.87% (C)	+6.03 points	Yes (*p*<.01)
Middle Grades Science	36.19% (F)	66.59% (D)	+30.40 points	Yes (*p*<.000)

**Note*. Pre- and Post-test letter grades are also provided in the table based upon a grading scale where A=90-100%, B=80-89%, C=70-79%, D=60-69%, F=59% and below.

IV. FINDINGS OVERALL FOR THEC STEM PD INVESTMENT – RESEARCH QUESTION 2

TEACHER SURVEY FINDINGS

An examination of the surveys that participants completed pre- and post-program revealed findings related to teacher opinions, frequency of use in instructional practices, student activities, instructional influences, teacher preparedness, principal perceptions, parental support, and professional development experiences. There were 452 participants that completed the pre and post survey.

Teacher Opinions Related to STEM Teaching

This construct examined teacher opinions regarding implementing effective STEM instructional strategies and access to associated resources necessary for doing so. A 10-item self-reported level of agreement construct, designed on a 5-point Strongly Disagree – Strongly Agree scale, evaluated teacher opinions. Overall teacher responses on this scale could range from 10-50. The THEC STEM PD participants demonstrated statistically significant improvement in opinions toward teaching mathematics/science from pre- to post-survey administration regardless of the PD program, $F(1) = 19.77$, $p < .000$. Additionally, there was a significant difference between groups with mathematics teachers having better attitudes at pre- and post-survey administration, $F(1) = 11.13$, $p < .000$. However, the difference was nominal, with mathematics teachers starting and finishing approximately 2 points higher than science teachers, who experienced similar growth.

Teacher attitudes significantly increased in agreement in areas such as feeling supported to try new teaching ideas, cohesion of school-wide teaching vision, and cooperation by sharing materials, and support by local agencies. Agreement with resource issues (i.e., time and computer access) was unchanged and remained relatively low (less than 50 percent agreement at pre/post). Enjoyment for teaching science/math agreement did not change, however, because it was extremely high at the pre-survey (93 percent) and remained similarly high at post-survey (93 percent).

Teacher Perceived Importance Related to STEM Teaching

This construct examined teacher-attributed importance of various use of instructional strategies, which are effective for STEM education. Thirteen items measured on a Not Important – Very Important scale assessed teacher importance. Overall teacher responses on this scale could range from 13-52. THEC STEM PD participants demonstrated statistically significant improvement in reported importance of use of effective mathematics/science instructional strategies from pre- to post-survey administration regardless of the PD program, $F(1) = .618$, $p = .450$. The difference between content areas was significantly different with science teachers remaining higher than math teachers at both pre- and post-survey, $F(1) = 4.85$, $p < .05$.

Teachers significantly increased their reported perceived importance of strategies in areas such as determining how to develop lessons (i.e., concrete experiences shared before abstract). Teachers' perceived importance of how students should engage with science content also significantly increased (i.e., inquiry-oriented activities, project/lab/research reports, and computer use). The perceived importance of developing students' conceptual understanding of the content and having students participate in appropriate hands-on activities both remained unchanged and high. More than 80 percent of the teachers surveyed reported these items were fairly or very important at both pre- and post-survey administration.

Instructional Influences

This construct examined the external influences teachers experienced that impacted whether or not they chose to use effective STEM pedagogy. Teacher perceived instructional influences were evaluated with 12 items on a 3-point scale assessing degree to which a factor inhibits or encourages effective instruction. Overall scores could range from 12-36. THEC STEM PD participants experienced statistically significant growth in this area – which means their impression of the influence of negative external pressures on their decisions to use effective pedagogy decreased from the beginning to end of program participation, $F(1) = 3.08$, $p < .05$. There was a statistically significant difference between groups based on content focus, $F(1) = 5.74$, $p < .05$. Teachers in the math PD programs averaged approximately a 2-point increase (on the 5-point scale), while teachers in the science PD programs averaged approximately only a .5-point increase. In practical terms, teachers from the science PD programs on average reported they had mixed feelings on whether the items inhibited or encouraged effective instruction, while teachers in the math PD programs reported this pre-survey but shifted closer to believing the items encouraged effective instruction at the end of the professional development program.

In all instances except for one which stayed similar from pre/post (state/district testing policies/practices), teachers perceptions of factors influencing their instruction became more positive as they shifted to feeling the factors encouraged effective instruction at a greater rate. However, at the post-survey more than 50 percent of the respondents reported that factors such as funds, time, and public attitudes still inhibited effective instruction.

Teacher Preparedness

This construct examined teacher perceived preparedness for teaching STEM content and use and delivery of effective STEM pedagogy. Teacher preparedness was assessed through 19 items on a 4-point scale (Not Prepared, Somewhat Prepared, Fairly Well Prepared, and Very Well Prepared) examining participants' self-reported sense of preparedness for STEM teaching in regard to content and pedagogical skills. Scores could range from 19-76. THEC STEM PD participants demonstrated statistically significant increases in preparedness to use various effective mathematics/science instructional strategies from beginning to end of program, $F(1) = 131.28$, $p < .000$. Additionally, there was no statistically significant difference between groups based on content focus (mathematics/science), $F(1) = 0.70$, $p = .446$. Overall, teachers increased from feeling Somewhat Prepared to Fairly Well Prepared and Very Well Prepared. Teachers reported feeling more prepared to do things such as provide concrete experiences before abstract concepts, develop student conceptual understanding, engage students in inquiry-oriented activities, and lead a class using investigative strategies.

Frequency of Use of Effective Pedagogy

Teacher frequency of use of effective pedagogy was determined through participant self-reported data on 14 survey items on a 5-point scale (Never, Rarely, Sometimes, Often, and Almost All Lessons). Overall scores could range from 14-70. THEC STEM PD participants reported statistically significant gains in use of effective pedagogy from pre- to post-survey, $F(1) = 10.77$, $p < .001$. Additionally, there was no statistically significant difference between groups based on content focus, $F(1) = 9.98$, $p < .001$.

All reported use of instructional practices increased from pre- to post-survey with the exception of two (introduce content through formal presentations and assign homework). Most of these practices saw a significantly positive shift, with nearly all being near or more than 75 percent of teachers indicating Frequently Used, except for the item regarding comment on reflections. This item still saw a positive shift but approximately 40 percent of teachers reporting doing this frequently at post-survey.

Student Activities

This construct examined the use of effective STEM instructional activities with student as the focus. The use of cooperative groups, student generated questions for investigation, communicating findings with others, use of technology, and other student-centered practices were the context for this construct. Student Activities employed in the classroom were evaluated with 20 items on a 5-point scale assessing how often a teacher has students engage in various effective instructional activities (Never, Rarely, Sometimes, Often, and Almost All Lessons). Overall scores could range from 20-100. A statistically significant increase in use of effective student activities was found for THEC STEM PD program participants, regardless of PD program, from pre- to post-survey, $F(1) = 31.36$, $p < .000$. Additionally, there was no statistically significant difference between groups based on content focus, $F(1) = 1.74$, $p = .232$.

Teachers increased their use of effective student instructional practices from Sometimes to Between Sometimes and Often. A majority of items in this section were reported as significantly increasing from pre- to post-survey (approximately 75% of items). Two items that did not significantly increase shifted down slightly (review homework assignments and read other science/math-related materials) and one other remained below 50% agreement (read from textbook) at both pre- and post-survey.

Parental Support

This construct examined the role of parents in STEM teachers' classrooms who participated in the THEC STEM PD programs. Parental Support was evaluated by six items on a 4-point scale assessing how many parents assist with different activities in the classroom (None, A Few, About Half, and About All). Overall scores could range from 6-24. A statistically significant increase in Parental Support was found for THEC STEM PD program participants regardless of PD program from pre- to post-survey, $F(1) = 6.08$, $p < .01$. Additionally, there was not a statistically significant difference between groups based on content focus, $F(1) = 2.51$, $p > .05$. Most items showed teachers felt unsupported by parents both before and after, with a vast majority of teachers selecting None or Few parents helping with all activities. Areas of significant growth included parents volunteering (2 percent to 6 percent), donating money or materials for the class (8 percent to 11 percent), and attending parent-teacher conferences (33 percent to 37 percent).

Principal Support

This construct examined the role of administrative support in the teaching of STEM. Principal Support was evaluated by nine items on a 5-point scale assessing the degree of agreement a teacher feels with the statements (SD, Disagree, No Opinion, Agree, and SA). Overall scores could range from 9-45. A statistically significant increase in Principal Support was found regardless of PD program from pre- to post-survey, $F(1) = 5.12$, $p < .01$. Additionally, there was not a statistically significant difference between groups based on content focus, $F(1) = 1.37$, $p < .251$. On average, teachers increased approximately 2 points on the Principal Support scale moving from between No Opinion and Agree to averaging a response of Agree.

Three Principal Support items saw a significant shift from less to more agreement (e.g., providing materials/equipment for science/math, providing time for teachers to meet and share ideas, encouraging teachers to observe other science/math teachers). All other Principal Support areas were notable because they had high levels of agreement at both pre/post (70 – 90%).

Professional Development Experiences

This construct examined the experiences and impressions of the THEC STEM PD participants regarding the individual program they participated in. The baseline measure asked participants to reflect on their past experiences with PD. The final survey participants were asked to respond if their impressions of the value of PD had changed relative to their participation in the THEC STEM PD Experiences were evaluated using three items on a 5-point scale assessing the extent to which participation in the district-offered professional development had increased teachers' abilities (Not at All to A Great Extent). Overall scores could range from 3-15. A statistically significant increase in PD Experiences was found regardless of PD program from pre- to post-survey, $F(1) = 10.02$, $p < .01$. Additionally, there was a statistically significant difference between groups based on content focus, $F(1) = 9.59$, $p < .01$. On average, math teachers felt more positively about their PD Experiences at the post-survey than did teachers in the science programs. Regardless of program, the average increase in PD Experiences was approximately 15% points from pre- to post-survey. However, even with these significant increases agreement failed to reach an average of at least 50% on any item.

V. CONCLUDING OBSERVATIONS

The THEC STEM PD Program investment revealed substantial growth in STEM teacher quality across the state of Tennessee. In this section we will present some concluding observations and highlights of the evaluation report. Individual narratives for each program are included as appendices to this report.

IMPROVED PEDAGOGICAL SKILLS

THEC STEM PD program participants demonstrated significant growth in STEM pedagogical skills, as observed in participant-submitted digital recordings of their instruction. The ability of teachers to design effective STEM lesson increased from 2.26 to 2.49 on the 5-point scale. Teacher implementation of effective STEM instruction also increased significantly from 2.48 to 2.96. Additionally, participants were able to transform their learning environments and create classroom culture, which supports investigative STEM education (2.57 to 3.10).

Participants' self-reported data on administered pre- and post-surveys indicated significant growth overall in opinions related to their own preparedness to teach STEM, frequency of use of effective STEM pedagogy (e.g., cooperative groups, technology, connections between science/math), use of student-centered activities, and connecting learning to the real-world.

IMPROVED CONTENT KNOWLEDGE

Classroom observations of THEC STEM PD program participants also revealed significant growth in content knowledge delivered during instruction (2.70 at baseline to 3.26 at end of program). Further, this growth was also reflected in program-developed assessments of content knowledge. Analysis of overall program developed content assessment data for THEC STEM PD programs revealed statistically significant growth from pre- to post-test ($F(94) = 6.09$, $p < .000$).

IMPROVED OPINIONS

Teachers who attended THEC STEM PD programs exhibited improved attitudes toward the teaching of STEM, as well as more positive experiences with parent and principal support. Further, participants felt more supported by colleagues, valued the use of inquiry, technology, and collaborative learning. Importantly, most participants valued the PD experience.

PROGRAMS CONSIDERED BEST PRACTICE

An examination of the evaluation data at the program level for the 16 THEC STEM PD programs revealed several programs that had significant impact on transforming STEM teacher quality (pedagogical skills) and content knowledge. The programs that improved both content knowledge and teacher quality, which could be considered best practice in our opinion, are listed below. Programs are listed alphabetically, by round.

Round One Programs

1. Austin Peay University – Grades 3-5 Mathematics (Principal Investigators Assad and Wells)
2. East Tennessee State University – Grades 3-5 Science (Principal Investigators Tai and Ho)
3. East Tennessee State University – High School Chemistry (Principal Investigators Rhoton and Zhao)
4. Tennessee Technological University – Grades 3-5 Science (Principal Investigators Gore and Hunter)
5. Tennessee Technological University – High School Chemistry (Principal Investigators Rust and Stevens)

Round Two Programs

1. East Tennessee State University – High School Chemistry & Physics (Principal Investigators Rhoton and Zhao)
2. Lipscomb University – Grades 4-7 Mathematics (Principal Investigators Wells, Morel & Nelson)
3. Lipscomb University – High School Algebra (Principal Investigators Nelson and Thornthwaite)
4. Middle Tennessee State University – Grades 4-8 Mathematics and Science (Principal Investigators Kimmins and Winters)
5. Middle Tennessee State University – High School Mathematics (Principal Investigators Strayer and Brown)
6. Tennessee Technological University – Grades 3-6 Mathematics and Science (Principal Investigators Pardue and Howard)
7. Tennessee Technological University – High School Mathematics and Science (Principal Investigators Fidan and Baker)
8. Tennessee Technological University – K-2 Mathematics and Science (Principal Investigators Baker and Fromke)
9. Tennessee Technological University and Roane State Community College – High School Mathematics and Science (Principal Investigators Suters and Lee)
10. University of Tennessee at Chattanooga – Grades 4-7 Science (Principal Investigators Ingraham, Ellis, and Carver)
11. University of Tennessee at Martin – Grades 6-12 Science (Principal Investigators Cox and Withmer)

SUMMARY

This final report for THEC on the STEM PD program has revealed teacher participation in the THEC STEM programs has resulted in overall growth in science and mathematics teacher effectiveness and attitudes in the state of Tennessee. At an individual program level, findings revealed many THEC funded programs also had significant impact on participants in all areas.

REFERENCES

Desimone, L.M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*(3), 181-199.

Johnson, C.C., Kahle, J.B., & Fargo, J. (2007a) A study of sustained, whole-school, professional development on student achievement in science. *Journal of Research in Science Teaching. 44*(6), 775-786.

Johnson, C.C., Kahle, J.B., & Fargo, J. (2007b) Effective teaching results in increased science achievement for all students. *Science Education. 91*(3), 371-383.

Johnson, C.C. & Fargo, J.D. (2010). Urban school reform through transformative professional development: Impact on teacher change and student learning of science. *Urban Education, 45*(1), 4-29.

Loucks-Horsley, S., Love, N., Stiles, K. E., Mundry, S., & Hewson, P. W. (2003) *Designing professional development for teachers of mathematics and science.* Thousand Oaks, CA: Corwin Press.

Onwuegbuzie, Anthony J. (2000, November). Expanding the framework of internal and external validity in quantitative research. Paper presented at the annual meeting of the Association for the Advancement of Educational Research, Ponte Vedra, FL.

Putnam R. & Borko, H. (1997). Teacher learning: Implications of new views of cognition. In B.J. Biddle, T.L. Good, & I.F. Goodston (Eds.), *The international handbook of teachers and teaching* (pp.1223-1296). Dordrecht, The Netherlands: Kluwer.