4-2019

# Measuring and Modelling How and When Effective Science Teaching Occurs

Elizabeth B. Lewis
*University of Nebraska-Lincoln*, elewis3@unl.edu

Lyrica L. Lucas
*University of Nebraska-Lincoln*, lyricalucas@huskers.unl.edu

Amy N. Tankersley
*University of Nebraska-Lincoln*, amntank@gmail.com

Elizabeth F. Hasseler
*University of Nebraska - Lincoln*, ehasseler@huskers.unl.edu

Brandon A. Helding
*Boulder Learning, Inc.*, b.a.helding@gmail.com

## Measuring and Modelling How and When Effective Science Teaching Occurs

Elizabeth Lewis, Lyrica Lucas, Amy Tankersley, Elizabeth Hasseler
University of Nebraska-Lincoln

Brandon Helding
Boulder Learning, Inc.

**Abstract**

      With new national science education standards, we must understand how to prepare science teachers capable of advancing reform initiatives. In a 3-year longitudinal study we adopted a multi-method approach to investigate beginning science teachers' instructional practices. We analyzed transcripts, administered a teaching self-efficacy survey, observed science lessons, and documented weeks of lessons. Using this large dataset, we posed research questions about the use of NGSS scientific practices in teachers' science lessons (Paper #1) and teacher- and student-level characteristics as it relates to teachers' use of inquiry in the classroom (Paper #2). In order to expand our coding capability of science teaching data for use in our structural equation modelling efforts (Paper #4) we also completed an initial validation of the DiISC instrument (Paper #3). Findings included: (a) differential use of scientific practices by physical and life science teachers in their lessons; (b) beginning teachers had lower levels of assessment use and there was little evidence to suggest that assessment varied greatly by classroom diversity; (c) evidence for the validity of the DiISC with factor analyses, correlations with the EQUIP instrument, and think-aloud and semi-structured interviews with DiISC raters; and (d) an SEM showed master's level teachers exhibited greater initial use of inquiry-based instruction and growth over time than undergraduate certified teachers with many contributing factors.

**Corresponding author:**
Dr. Elizabeth Lewis
Associate Professor, Science Education
University of Nebraska-Lincoln
Email: elewis3@unl.edu

**Introduction**

With new national science education standards, the *Next Generation Science Standards*

(NGSS), we must understand how to prepare science teachers capable of advancing science

education reform. In this related paperset we adopted a multi-method approach to investigate

beginning science teachers' instructional practices in a 3-year longitudinal study. We analyzed

transcripts, administered a teaching self-efficacy survey, observed science lessons, and documented

a week's worth of lessons including the observed lesson. Using this large dataset, we focused on

research questions about the use of NGSS scientific practices in teachers' science lessons (Paper #1)

and classroom diversity as it relates to teachers' use of assessment (Paper #2). In order to expand

our coding capability of science teaching data for use in our structural equation modelling efforts

(Paper #4) we also completed an initial validation of the DiISC instrument (Paper #3).

Findings from this study include: (a) differential use of scientific practices by physical and

life science teachers in their lessons; (b) beginning teachers had lower levels of assessment use and

there was little evidence to suggest that their assessment varied greatly by classroom diversity; (c)

the DiISC can be validated with paired coding along with the EQUIP instrument; and (d) an SEM

showed master's level teachers exhibited greater initial use of inquiry-based instruction and growth

over time than undergraduate certified teachers with many contributing factors.

**Overall Conceptual Framework**

In recent work we developed a growth model of teachers' learning and mediating factors that

may support or impede teachers' enactment of inquiry-based teaching (Figure 1, from Lewis, Rivero,

Musson, Lucas, Tankersley, & Helding, 2019; Lewis, Rivero, Lucas, Tankersley, & Helding, 2018).

Elements of the teacher preparation program are contained in the upper bracketed area of the figure

and the mediating factors are in the middle box with the arrows leading to reformed-based science teaching practices.
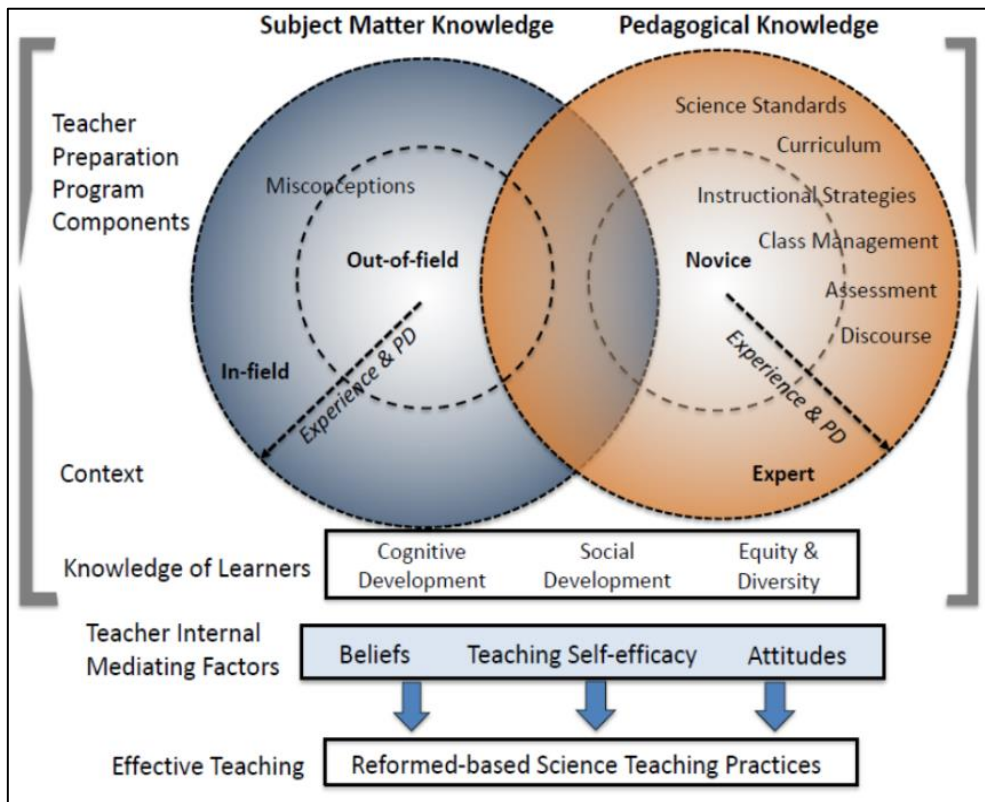


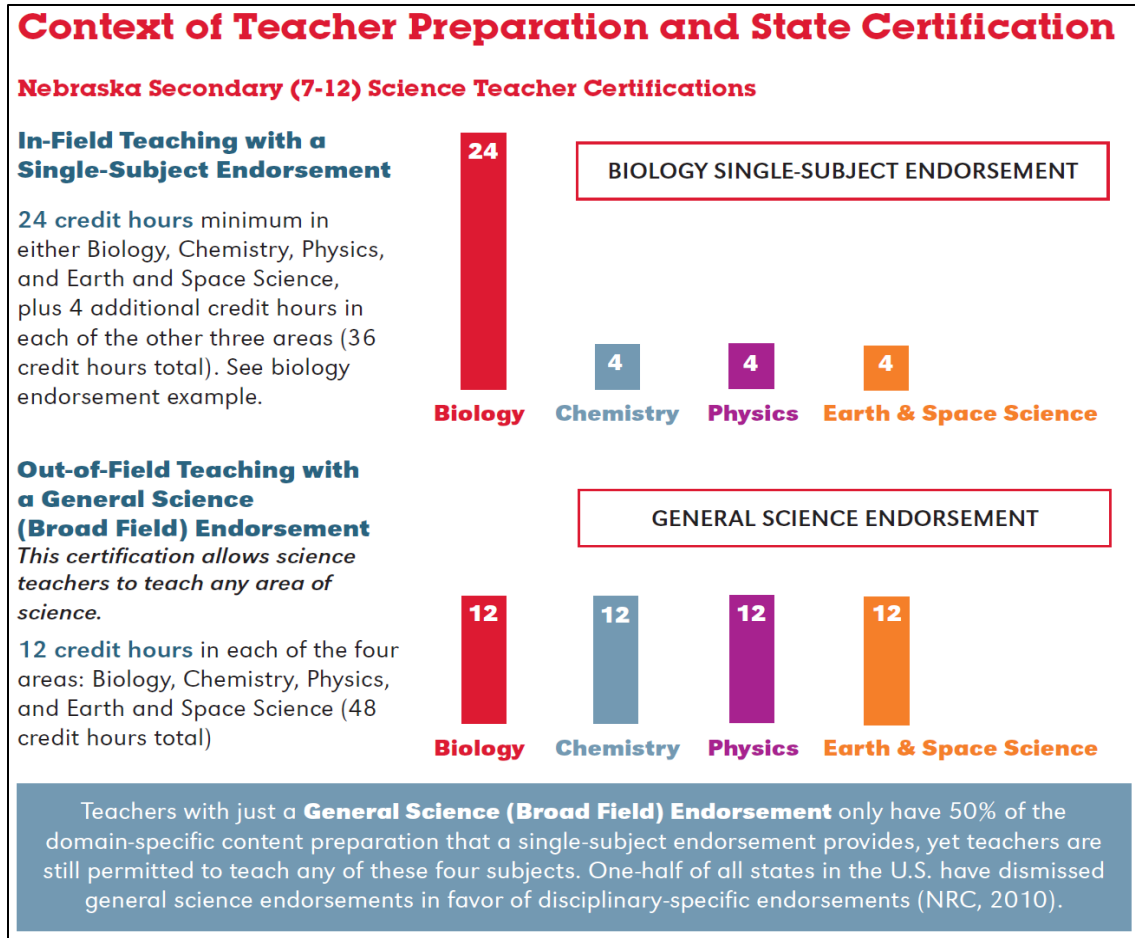*Figure 1.* Conceptual framework (Lewis, et al, 2019; Lewis, et al, 2018).

*Figure 2.* Single-subject and general science endorsements and their relative relationship to in-field and out-of-field teaching.

**Paper #1: Secondary Science Teachers Use of NGSS Science Practice in Teachers'
Science Lessons**

Amy Tankersley
University of Nebraska-Lincoln

## Introduction

The *Next Generation Science Standards* (NGSS) combines content standards, crosscutting concepts, and science and engineering practices into a cohesive framework for learning science. Scientific practices (SP) are a key facet of NGSS and are designed to help transform K-12 science from didactic learning to mirror the work of scientists, with a focus not only on learning the content but also the nature and practice of science (NGSS Lead States, 2013). Effective use of NGSS will require teachers to be able to plan, enact, and reflect on their lessons and support students in three-dimensional learning through NGSS (Schneider & Plasman, 2011). Future science assessments need to integrate relevant science practices they need to be able to integrate student's assessment of their knowledge of the core ideas together with science and engineering practices (NRC, 2010). To be able to support learning, teachers will need to be able to transform their teaching practices and develop effective assessments we will need to change the manner in which we prepare new teachers whose ideas may ideas about teaching and learning may not necessarily align with NGSS (Crawford, 2014; Bybee, 2014).

To diagnose problems in and revise teacher education we will need to first understand what science practices are being used and how teachers are using those practices in the science classroom. Despite the relative newness of the science practices, there has been some research into how teachers use science practices in the secondary classroom. Multiple factors are involved in the use of scientific practices, and teachers tend to focus on some of the practices more than others (Carpenter et al., 2015; Morales, 2016). In their early years of teaching second career science

teachers were influenced by their prior experiences when planning and implementing science practices into their lesson plans (Antink-Meyer & Brown, 2017). There is also some evidence that some of the practices are linked and that often teachers use analyzing and interpreting data, constructing explanations, and obtaining and evaluating information in conjunction with one another (Brownstein & Hovarth,2016). Despite a few prior studies, there is still little information on how students engage in practices in the classroom (Carpenter et al., 2015). Many of the studies that have been done using lesson plan analysis or small-scale studies with few teachers and there is a real need to analyze large scale lesson data to determine what practices are being used in the secondary science classroom and the factors that influence the use of those practices.

## Literature Review

### The Evolution from Inquiry to Science Practices

In the early 1990s, science education underwent a revolution where teachers began to focus on the quality of students' sense-making process (Taylor, 2014). Constructivist techniques see reality in the mind of the knower and people construct knowledge based on his or her perceptions (Jonassen, 1991). Constructivism has been a major emphasis in teacher education programs for more than 20 years (Ford, 2015). Despite this concentration on constructivism, new teachers are still more likely to use instructivist techniques in the classroom (Dalgarno and Colgan, 2007). The continued emphasis on instructivist techniques makes student learning in science a passive endeavor in which the teacher is the center of most lessons (Sizer, 1984). Many schools have resisted the change to inquiry-based science instruction and continue to use memorization without attention to coherent thinking or an appreciation for scientific explanations (Woodruff & Meyer, 1997). One of the reasons for this emphasis on instructivist practices is the continued belief that constructivist practices require students to use cognitive abilities they are not ready for yet (Kirschner et al.,

2006).  Those who champion this argument contend that novice and intermediate learners need significant support in learning that cannot be gained through constructivist learning (Kirchner et al., 2006).  Even if teachers are excited about constructivism and want to use student-centered techniques, they are often nervous at the actual implementation of inquiry techniques and relinquishing control of the classroom required when facilitating more student-centered learning (Porcaro, 2011).   Constructivist learning occurs most effectively in context and teachers should create environments where learning is relevant (Jonassen, 1991). Transitioning from instructor to facilitator requires different skills than many pre-service teachers have seen in their schooling (Porcaro, 2011).  Early career teachers have less developed views on education and may have beliefs less resistant to change than experienced teachers (Luft et al., 2011).  This makes preservice and new teachers a population that can be targeted for professional development in hopes of shifting their teaching practices to better fit a constructivist model.

Scientific literacy parallels general literacy in its requirement of abilities and background understandings to infer and interpret meaning from text, talk, and other modes of representation to build new interpretations (Cavagnetto,2010).  For teachers to be able to increase students' scientific literacy they will need proper teacher education and professional development.  Teachers will need to be able to improve pre-service teachers' ability to incorporating STEM concepts, prompting students to ask their own questions, developing and refining their questions and experimental design, and disseminating results to peers to the larger scientific community (French & Burrows, 2018).  Many teachers report that their primary goal for their students is to gain science skills such as critical thinking and problem-solving strategies, but few report their skills goals specifically to science (Sandoval & Kawasaki, 2016).

The introduction of NGSS and science practices into curriculum promises to increase science literacy beyond inquiry by integrating student learning of content knowledge with engagement in the practices of science.  In NGSS, each performance expectation combines content with relevant science and engineering practices and crosscutting concepts (NGSS Lead States, 2013).  To implement NGSS effectively, students will need to learn science by actively engaging in the practices of science including conducting investigations, sharing ideas with peers, engaging in specialized ways of talking and writing, mechanical, mathematical, and computer-based modeling, and development of representations of phenomena (NRC, 2007). The transition to NGSS and science practices will require teachers to provide instruction, lessons, activities, and assessments that connect and focus on ideas of and about science and engineering practices which are essential to helping students develop a deep understanding of science across disciplines (Kloser, 2014). To provide that instruction teacher educators will need to revise teacher education to facilitate the transformation of science education and science teachers (Crawford, 2014).  In order for that transformation to occur teacher educators will need to understand the current state of science education and the science practices used in the classroom.

**Research on the Use of Science and Engineering Practices**

As more states, schools, and districts move toward NGSS and three-dimensional science learning we are developing a growing body of literature on the science and engineering practices but there is still much we need to learn.  We have some evidence that teachers may better understand and implement some practices over others (Brownstein & Hovarth, 2016; French & Burrows, 2018).  One study of science teachers in at a professional development designed to support understanding and use of science and engineering practices found that teachers may already be comfortable with some practices like analyzing data and further professional development in

those areas may not be needed (French & Burrows, 2018). Further support for the high use of analyzing and interpreting data by teachers comes from a survey that reported that 59% of teachers reported using analyzing and interpreting data either very often or almost always but (Drew & Thomas, 2018).

Communication is a big part of the science practices and the ability of students to communicate their understanding of science. Effective classroom discourse via NGSS should be modeled on that of scientists which involves sharing multiple interpretations of phenomena, collaboration, and collaborative discourse where students share their ideas and request clarifications from their classmates (Amin, Smith, & Wiser, 2014). NGSS also puts greater emphasis on argumentation and modeling that allows students to go beyond inquiry by interpreting and evaluating data to develop arguments, explanations, and models (Crawford, 2014). Teachers surveyed about their use of science practices 44% of the 343 teachers surveyed reported that they had students construct explanations very often or frequently but most reported only using argumentation from evidence in the classroom only occasionally or lesson frequently (Drew & Thomas, 2018). Scientific argumentation is a key practice, but it is not used much in the science classroom and there needs to be research, attention, and planning to successfully engage science students (Antink-Meyer & Brown, 2017; Brownstein & Hovarth, 2016). Even when attempted students often struggle with to produce justification for claims, generate complex explanations, and develop coherence linking ideas together focusing instead on the claims and assertions (Kelly, 2014). The lack of use of argumentation is troubling because of the central role it plays in students' ability to use evidence to create arguments to science literacy (NGSS Lead States, 2013).

Modeling and using mathematical and computational reasoning are another set of science practices that are important to science literacy (NGSS lead states, 2013). Research on models and

modeling highlight the role of analogical restructuring and strategic recruiting of intuitive models in the process of conceptual change (Amin, Smith, & Wiser, 2014). For engineers, system modeling is essential to developing complex technologies and helps students understand the relationships between micro and macro scales and provide opportunities for students to reason scientifically (Cunningham & Carlsen, 2014). Mathematical and computational thinking is another form of modeling that allows students to represent physical variables and their relationships and thus make quantitative predictions (NGSS Lead States, 2013). Science can provide concrete examples of abstract mathematical ideas and math can provide ways to quantify and explain science relationships and thus deepened science knowledge (Czerniak & Johnson, 2014).

Along with research on the use of practices is literature on possible barriers to effective integration of science practices. Teachers often use science practices to reinforce concepts, engage with the content, learn the scientific method, or assess student understanding but rarely discuss using the practices as a holistic part of learning science (Sandoval & Kawasaki, 2016). This lack of an integrated view of science practices can limit their use in the classroom or the effectiveness when they are used. Pre-service teachers especially may not completely understand the practices and when surveyed often expressed confusion over practices like modeling and even when they could describe the practices the descriptions often did not entirely fit the full description of the practice as outlined in NGSS (Carpenter et. al, 2015). One study of 26 teachers in a professional development found that when discussing examples of practices integrated into their lessons most often described activities related to the scientific practices but not actually teaching core concepts through student engagement in scientific practices which highlights a need to create professional development that provides support to create lessons that use practices in a meaningful and integrated way (Sandoval & Kawasaki, 2016). Understanding what practices are being used and who is using those practices

may help us evaluate different groups of teachers and determine what barriers limit teacher ability

to use science practices in the secondary science classroom.

## Research Questions and Methods

Our research questions for this study were: *What and how often are science and engineering*

*practices used in the secondary science classroom? How do teacher and classroom factors*

*influence the use of science and engineering practices in the classroom?*

### Participants and Setting

This data is part of a larger set of data where we observed, interviewed, and collected data

on secondary science teachers who were graduates of the same Midwestern University.  For this

study analyzed data from 55 of the teachers over a 2.5-year period collecting 514 weeks of self-

reported classroom instructional practices in secondary science classrooms.  All participants had a

series of 2 methods courses. The first course was an overview of science teaching and the theory

and practices related to inquiry and NGSS.  The second course deepened the pre-service teachers of

science education through developing an understanding of curriculum and assessment.  All

participants also had more than 500 hours in the classroom during pre-service where they had the

opportunity to implement the instruction from their methods courses supported by cooperating

teachers and a university supervisor.  All participants were in their 1-7 years of teaching and many

came from two large districts around the University that were in the process of either piloting or

implementing standards adapted from NGSS and thus were expected to use the science and

engineering practices.  Of the 55 participants, 21 (38%) graduated with a bachelor's degree in

secondary science education and 34 (62%) with a Master of Arts in education from the same

university (Table 1).  Most of our lessons were taught by teachers who taught in-field. In our

Midwestern state, in-field teachers have to have a certification for teaching the subject. For our

purposes, we categorized in-field teachers as having at least 24 credit hours in the subject with a

state single subject endorsement for high school classes and at least 12 credit hours in the subject

(broad field endorsement) for teaching middle school.

Table 2. *Summary of participating teachers by program membership*

| Program | Number of Teachers (%) | Number of Lessons (%) | Level (%) | Lesson Subject Area (%) | In field/Out of field (%) | Mean SEP Used in a week (SD) |
|---|---|---|---|---|---|---|
| Undergraduate | 21 (38%) | 179 (35%) | Middle School = 67 (13%)<br>High School = 112 (22%) | Physical Sciences = 65 (13%)<br>Life Sciences =78 (15%0<br>Earth and Space Science = 35 (7%)<br>Other =1(.2%) | In Field = 127 (25%)<br>Out of Field= 52 (10%) | 5.12 (3.73) |
| MAT | 34 (62%) | 335 (65%) | Middle School = 77 (15%)<br>High School = 258 (50%) | Physical Sciences= 137 (27%)<br>Life Sciences = 147 (28%)<br>Earth Sciences = 48 (9%)<br>Other =3 (.6%) | In field= 255 (50%)<br>Out of field = 79 (15%) | 5.87 (4.22) |
| Total | 55 | 514 | Middle School = 144 (28%)<br>High School = 370 (72%) | Physical Sciences = 202 (39%)<br>Life Sciences = 225 (44%)<br>Earth and Space Science = 83 (16%)<br>Other = 4 (.8%) | In field = 382(74%)<br>Out of Field = 131 (25%) | 5.64 (4.11) |

**Data Collection**

The researchers interviewed each participant 4-6 times per year for each year they

participated in the study.  Teachers were interviewed asking them to summarize 4-5 lessons around

a class that was observed by one of the researchers.   To standardize the class time of the lessons

gathered we collected 5 days of lessons from schools that were on a "regular" 40 to 50-minute

schedule and 4 days of lessons from block schedules, which ranged from 85-120 minutes depending

on the school schedule. During each interview, the teachers were asked to summarize 3-4 lessons

previously to gather approximately a week's worth of data.  The researchers observed one of the

lessons in the summarized time period allowing us to better understand the instructional practices in

the classroom and ask leading questions to get the full range of practices that were used by the

teacher.  After that interview, we coded each day summarized by the teacher with a 1 or 0 to

indicate the use of the scientific practices as well as recording a brief qualitative description of the

practices used during each class period.  The data ($n$=514 weeks) was then averaged per week to

obtain the average frequency of each practice and overall SEP practice use.  Data were then sorted

according to the program, subject, level (MS vs. HS), and in-field vs. out-of-field teaching.

**Data Analysis**

For each science and engineering practice, we calculated the percentage of days in each

week that the teacher described a lesson activity where a SEP was involved.  We also calculated the

number of weeks that at least one practice was used during the week and the number of practices

used each week by the teachers.  To account for differing lengths of class periods and standardized

the data all other analysis was done and the unit of the week and not the individual lesson.

To determine possible teacher, school, and classroom factors that might influence the use of

science practices. From the possible factors that might impact the science practices we selected

program, subject matter of the lesson, level (middle school (MS) or high school (HS), in or out-of-

field teaching as possible factors that might influence the use of science and engineering practices.

We reviewed each teacher and school and  first calculated descriptive statistics, finding the mean

and standard deviation for each practice use by: school level (middle school and high school), in-

field and out-of-field teaching, subject category (i.e., physical science, life science, Earth and space

science), years of teaching experience (i.e., 0-3 years or 4-7 years), and by program (undergraduate

or MAT teachers).  We then calculated the mean and standard deviation for the total number of

practices used in each group as well as the mean and standard deviation for the percentage of weeks

that at least one practice was used by the teacher.  We followed up the descriptive statistics with a

MANOVA type I sum of squares followed by a Tukey's post hoc test when there were more than

two categories of participants.  Results were reported as significant for the MANOVA if the p-value

was less than .05 and significant between-group effects if the p-value was less than .025.

## Results

In this section, we briefly report our participants' overall use of science practices as well as

factors that influence teachers use of science and engineering practices in the secondary science

classroom.  We start by reporting the most used and least used practices for all of our participants

followed by the factors that significantly impacted the use of science practices.  For each factor that

was determined to be significant, we used between group effects to report which science practices

varied significantly by each factor.

Overall our teachers used at least one science practice for 96% of the weeks surveyed with a

standard deviation of 20%.  This averaged to approximately 15% of the week's lessons using one or

more of the science practices.  Analyzing and interpreting data was the most commonly used SEP

with an average of 27%(SD=26%) of the lessons per week followed by mathematical and

computational thinking 25% (SD=34%) and Asking questions and defining problems at

21%(SD=33%).  Engaging in argumentation from evidence at 2% (SD=8%) and constructing

explanations and designing solutions 8% (SD=19%) were the least used by our participants and

were recounted in only a few of the lessons surveyed.

Next, when looking at the factors that influenced the use of science practices, we started

with program level.  Participants that were graduates of the MAT program used significantly more

science practices per week than teachers who graduated from the program with a bachelor's degree,

$F_{(8.503)} = 2.763$, p=.005; Wilk's $\Lambda$ = .958, $\eta^2$ = .042.  Table 2 shows that teachers who had a

Master's degree had their students engage in planning and carrying out investigations ($F(1,510) = 5.615$, p=.018, $\eta^2 = .011$), analyzing and interpreting data ($F(1,510) = 11.766$, p=.001, $\eta^2 = .023$), and using mathematics and computational thinking ($F(1,510) = 7.845$, p = .005, $\eta^2 - .015$) more frequently than participants who had a bachelor's degree.  The only science and engineering practice that was used more often by teachers who graduated from the undergraduate program was engaging in argumentation from evidence, which was used 2.6% of a week's worth of lessons on average by the participants from the undergraduate program and only 2.1% of the week on average by participants from the master's program, but the difference was not significant.

Table 2. *Use of Science and Engineering Practices by Program*

| Science and Engineering Practice | Undergraduates n=179 (SD) | MAT n=355 (SD) | Total n=514 (SD) |
|---|---|---|---|
| Asking Questions and Defining Problems | 21% (33%) | 22% (34%) | 21% (33%) |
| Developing and Using Models | 13% (20%) | 14% (24%) | 13% (22%) |
| Planning and Carrying out Investigations | 8% (17%) | 12% (23%) * | 11% (21%) |
| Analyzing and Interpreting Data | 22% (24%) | 30% (28%) * | 28% (26%) |
| Using Mathematics and Computational Thinking | 19% (29%) | 28% (36%) * | 25% (34%) |
| Constructing Explanations and Designing solutions | 7% (17%) | 9% (21%) | 8% (19%) |
| Engaging in Argumentation from Evidence | 3% (10%) | 2% (8%) | 2% (8%) |
| Obtaining, Evaluating, and Communicating Information | 14% (26%) | 16% (25%) | 15% (26%) |

*\* indicates a p<0.05*

There was also a significant difference in the average percentage of practices used per week by the subject matter of the lessons (40, 2186) = 7.079, p< .001; Wilk's $\Lambda$ = .588, $\eta^2 = .101$).  The subject matter of the lesson had a significant effect in the use of average use per week of developing

and using models (F(5,508)= 4.422, p =.001, $\eta^2$ = .042), planning and carrying out investigations (F(5,508) = 5.982, p <.001, $\eta^2$ = 056), analyzing and interpreting data (F(5,508) = 4.942, p <.001, $\eta^2$ = .046), using mathematics and computational models (F(5,508) = 36.591, p <.001, $\eta^2$ = .265), and constructing explanations and designing solutions (F(5,508) = 9.480, p<.001, $\eta^2$ = .08) (Figure 1). Physics teachers used science and engineering practices significantly more often than chemistry, physical science, biology, and Earth and space science teachers.  When looking at the use of developing and using models, planning and carrying out investigations, and constructing explanations and designing solutions physics teachers on average used those practices significantly more in a week than any other subject (P<.05).  Chemistry teachers used modeling significantly less than either physics or Earth science teachers (p<.05).  Chemistry teachers also used to analyze and interpreting data less often in their lessons than both physics and physical science teachers in our study (p<.05).  Biology and earth science teachers did not use any practice significantly more than other subjects. In fact, biology and earth science teachers lagged behind physics teachers in many of the practices and did not use any practice significantly more than any other subject.
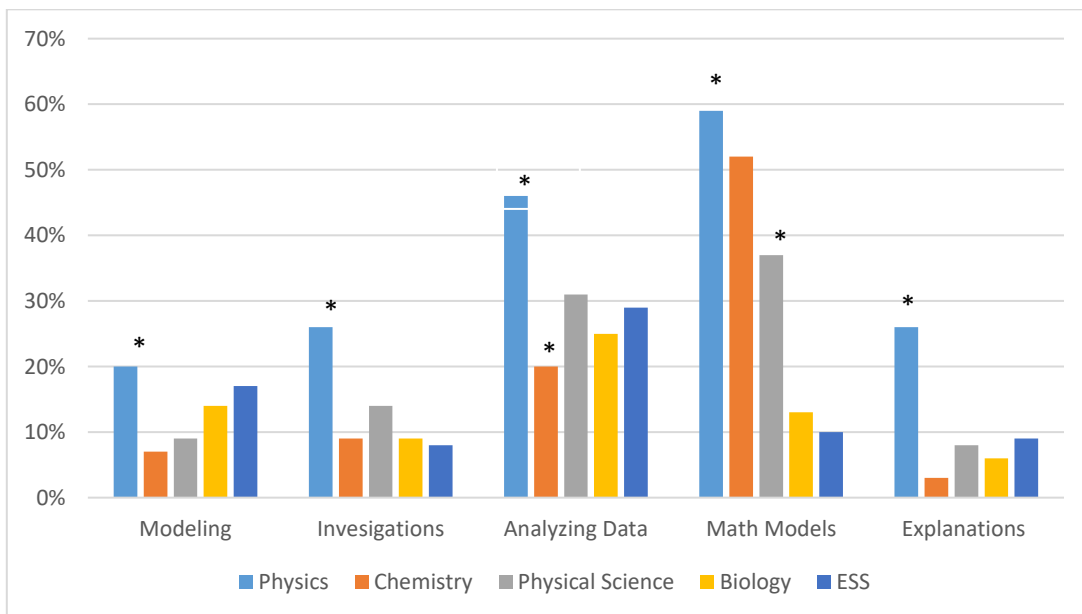


*Figure 1.* Use of science and engineering practices by lesson subject.

\* indicates a P< .05

Science teacher training consists of not only pedagogical training but training in the discipline they are planning to teach as well. Teachers need to have strong subject matter knowledge to understand and implement curricular content and goals in a way that addresses student needs (Darling-Hammond, 2016). Teachers who taught out of subject area found difficulty selecting resources and using those resources to help students gain knowledge and make links between concepts in the discipline (Childs & McNicholl, 2007). In this case we defined out of field teaching for middle schools as teachers who either do not have a single subject certification, which requires at least 24 college credit hours in the subject, in the lesson subject area or a do not have a broad field certification which requires at least 12 hours in lesson subject matter and the passing of a general science subject area tests. For high school lessons, we defined out of field teachers as not having at least 24 credit hours in the subject area of the lessons. For our participants there was no significant difference overall in the use of science and engineering practices for teachers that taught in-field and out-of-field (8,505) = 1.329, p=.226, Wilks' $\Lambda = .979$ ŋ2=.612). Because of the trends, we saw by the subject matter of the lessons we decided to disaggregate in and out-of-field teaching by the subject matter of the lesson. Through analysis of the influence of out of field teaching by lesson subject matter we found that that teachers who taught chemistry lessons out of field used significantly less science and engineering practices in their lessons than teachers who had at least 24 college credit hours of chemistry and thus were teaching in-field (F(8.47)=2.359,p=.032, Wilks $\Lambda = .713, \square 2 = .832$)). Specifically, chemistry teachers who taught in-field used planning and carrying out investigations (p=.001) and analyzing and interpreting data (p=.003) more often than teachers who did not have a single subject certification to teach chemistry (Figure 2).
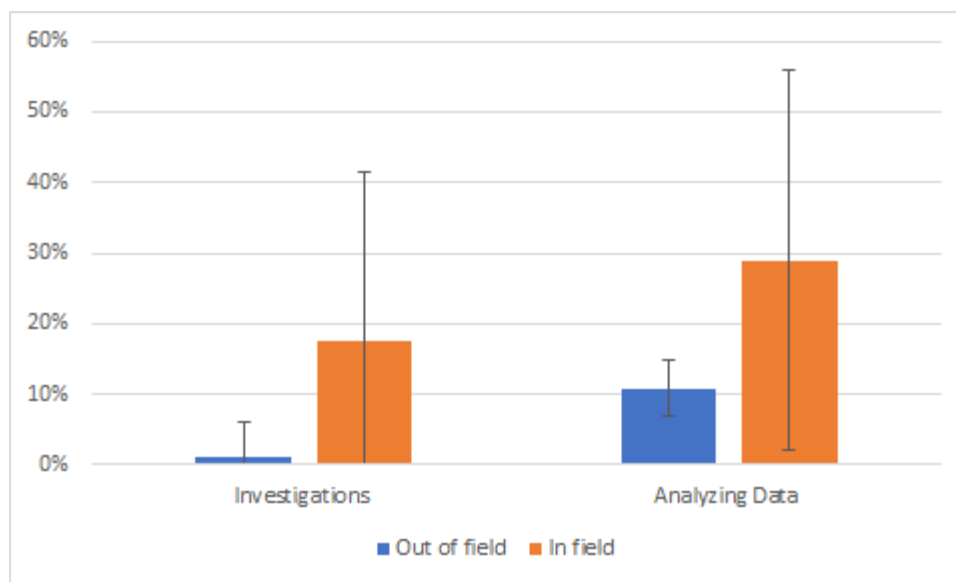
*Figure 2.* Use of Science Practices by In and Out-of-field Chemistry Teachers

In our study, 52% of the chemistry lessons were taught by teachers who were not certified to teach the subject and thus out of field teachers in chemistry impacts students access to the science and engineering practices during their chemistry courses. We were not able to properly examine the influence of out of field teaching on physics because we only had 38 total lessons of physics in our analysis and of those only 9 or 24% of the lesson were taught by teachers not certified in the subject. We were also not able to analyze the influence of out of field teaching on the use of science and engineering practices in life science lessons because only 2 out of the total 212 life science lesson or .9% were taught by teachers not certified in the subject.

Concentrated poverty affects students learning because schools with a high number of low SES students tend to have less experienced teachers and fewer learning resources, more limited curricula taught at less challenging levels and other factors that can significantly affect academic achievement (Darling-Hammond, 2013). Because of this we were curious about the impact on SES on the use of science practice in the secondary science classroom. For this analysis, we defined schools that were high needs schools as middle schools that had more than 50% of the students

qualifying for free and reduced lunch and high schools that have more than 40% of their student's

body qualifying for free and reduced lunch. Table 3 shows the number of lessons taught in high or

low socioeconomic schools and the average use per week of the science practices. According to our

analysis, teaching in a high needs did impact the use of science and engineering practices in the

classroom $F(16,1008) = 3.501$, p<.001, Wilks' $\Lambda = .897, \Box 2 = .053$. Teacher who taught in higher

socioeconomic schools had students engage in planning and carrying out investigations

$(F(2,511)=8.249, P<.001, \eta2 = .031$, analyzing and interpreting data $(F(2.511) = 3.751, p=.024,$

$\eta2 = .014$, and constructing explanations and designing solutions $(F(2.511) = 14.135, p<.001, \eta2 =$

*.002* on average more times per week than teachers who taught in schools with a higher proportion

of students that qualified for free and reduced lunch.



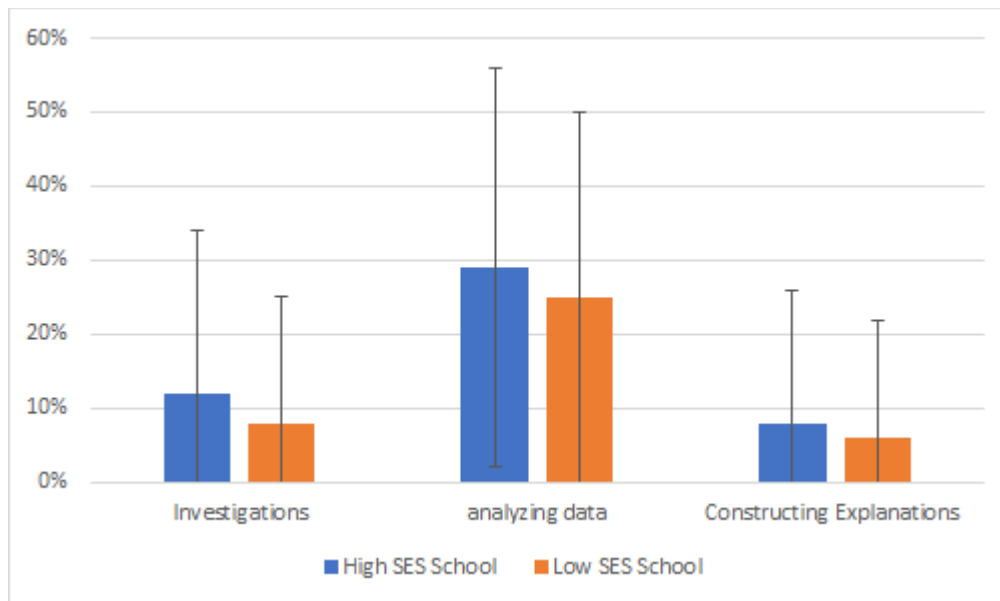Figure 3. Use of science and engineering practices in lessons conducted in high needs schools
* indicates p<.05

## Discussion

Our study builds on and extends prior research into the use of scientific practices in

the classroom. Teachers have been shown to some practice more than others (Brownstein &

Hovarth, 2016; French & Burrows, 2018) and in our analysis, we found that overall teachers were

more likely to integrate analyzing and interpreting data, using mathematical and computational thinking, and asking questions and defining problems into their instructional practices. Practice that involved discourse was especially hard for our teachers and engaging in argumentation from evidence and constructing explanations and designing solutions were rarely used by our teachers. This has implications for the continued resistance of teachers to high-yield discourse strategies and the lack of higher order discourse in the science classroom.

Beyond the overall conclusions about science and engineering practices were found some teacher and school factors that influenced the use of science practices in the science classroom. Our teachers who had a master's degree averaged a higher percentage of practices per week than teachers from our undergraduate program. Prior research has indicated that career changers in science education bring some of the ideas and orientations from their original field of science study (Antink-Meyer & Brown, 2017). On average the teachers from the MAT program typically have a high number of science courses in their content area, take more 300 and 400 level courses than our undergraduates, and often have experience in science fields which may carry over to their instructional practices and ability to implement the science practices.

The subject matter of a lesson also impacted teachers' use of science practices with physics teachers using practices like modeling, investigation, analyzing and interpreting data, mathematical and computational thinking, and explanations more than biology, chemistry, and earth science teachers. Chemistry teachers low use of science practices may have been influenced by the high number of our participants teaching chemistry out-of-field. Chemistry teachers with at least 24 credit hours of chemistry coursework integrating more investigations and analyzing data than teacher without a single subject endorsement in chemistry. This supports prior work on the importance of subject matter knowledge to reform-based teaching practice and assertion that

chemistry subject matter knowledge influences their use of inquiry in the classroom (Lewis et al., 2018). Finally, our work points to the need for more research into the continued inequities of low SES schools and the need to find supports for teachers who are implementing NGSS and the science practices in high needs environments.

## Limitations and Future Directions

This data is self-report by teachers and why we strengthened the data with observations of teacher lesson plans we relied on teachers to summarize their lessons effectively. We also used a sample of convenience and thus did not have sufficient coverage in all categories to complete a full analysis of many of the factors that might influence the use of science practices. We also confined our participants to the alumni of one midwestern university and therefore cannot generalize our findings to teachers from other programs or contexts.

In the future we would like to perform a hierarchical linear model on our data to predict use of science practice by the secondary science teachers in our study. We would also plan to use qualitative analysis to analyze researcher observations and teacher interviews to better understand not only what practices are being used and by whom, but also how those practices are used in the secondary science classroom. the ultimate hope for this work is to find gaps in teachers understanding and use of the practice and develop professional develop that will help teachers all of the science practices consistently and effectively.

**Paper #2: Factors Impacting Teachers' Use of Inquiry in the Science Classroom**

Elizabeth F. Hasseler

University of Nebraska-Lincoln, Department of Teaching, Learning, & Teacher Education

**Abstract**

This study investigates student- and teacher-level factors that impact the use of inquiry in the classroom by teachers who were prepared through one of two programs at the same university. Student-level factors that were used were their gender and racial demographics and grade level (MS/HS). Teacher-level factors included years of experience and membership in either a BA or MA leading to certification teacher preparation program. We used over 650 lesson observations that were generated between 2015 and 2018 and were coded using the EQUIP rubric. Multiple regression was used to investigate the EQUIP scores. We found that the diversity of the students was not correlated with use of inquiry in the classroom, but that membership in the teacher preparation program was a significant factor.

**Introduction**

Through my observations of local middle and high schools, I have seen science classrooms of varying student demographics, with respect to gender and race. I have also noticed that upper-level science classes tend to be of similar demographics, where most students are White, regardless of the demographics of the school. This observation suggests many questions regarding the program enrollment mechanisms at the schools; as well as, whether the teaching practices in the more diverse classrooms are different than those of a less diverse classroom. If the teaching practices can be predicted based on the demographics of the classroom, then maybe that could also impact students' success and whether they move onto more advanced science courses.

In this study, I will be delving into the following question: Is the use of inquiry by teachers predicted by the demographics of the classroom? In the future, I would like to take this question further and ask whether the use of inquiry is predicted by type and level of the science course.  I chose to research inquiry in the classroom because of the numerous benefits to all students.

## Conceptual Framework

The conceptual framework presented in the introduction to the paperset was developed by our research team in an earlier study (Lewis, et al, 2018). For this paper's study, we focused on teachers' pedagogical knowledge from their teacher preparation program and years of teaching experience.  We are also focused on student factors of gender and racial diversity, as well as, their grade level, which was broken down into middle school or high school.  From the framework, all of these factors have a relationship with and contribute to reform-based science teaching practices.

## Literature Review

### Inquiry

Sociocultural learning theory has its roots in Vygotskyan tradition (Lemke, 2001).  Part of Vygotsky's theory analyzed the interplay between language and learning.  He "conceptualized development as the transformation of socially shared activities into internalized processes" (John-Steiner & Mahn, 2012).  Sociocultural theory looks at the questions of how personal identity and cultural values impact the science classroom.

The goal of integrating inquiry into the classroom is to help students to develop scientific reasoning (Chinn & Malhotra, 2002). Open inquiry in the classroom has many facets.  It includes students designing procedures, making hypotheses, discussing results, and  "linking of experience to activities, science concepts, and science principles" (Roth & Bowen, 1995). Inquiry-based

instruction has its roots in Dewey, who believed that education should be based on experiences in the classroom (Dewey, 1938).

The EQUIP rubric was developed to be able to assess the use of inquiry. It was developed as a tool for researchers and teachers to push inquiry beyond using student-centered activities to support "students critically and systematically engage in examining, interpreting and analyzing questions regarding the world around them and then communicate their findings, providing convincing arguments for their conclusions." It is broken down into four sections: discourse, instruction, assessment, and curriculum (Marshall, Smart, & Horton, 2010).

One instructional model that facilitates the development and use of inquiry in the classroom is the 5E model. 5E stands for engagement, exploration, explanation, elaboration, and evaluation. Through the engagement process, teachers begin the lesson or unit with an activity that hooks the students' interest and connects with their prior knowledge. Through exploration, students learn and discover new concepts and ideas through activities such as labs or simulations. Afterwards, is the explanation process where students show what they have discovered and learned through the prior processes. Elaboration is where the extension of students' learning is occurring through activities such as classroom discourse. Finally, the learning is evaluated. (Bybee et al., 2006).

Inquiry is used in the classroom in three different ways, "scientific inquiry" with respect to how scientists *do* science, "inquiry learning," referring to the students engaged roles in inquiry-based instruction, and "inquiry teaching" which refers to the pedagogical choices by the instructor. Through Crawford's chapter that focuses on inquiry in the classroom, she notes that inquiry goes beyond hands-on activities in the classroom. It involves *doing* science as a scientific practitioner. Inquiry involves students being able to also asking questions, designing experiments, developing

models argumentation from evidence, such as what is recommended through the Next Generation

Science Standards (Crawford, 2014).

**Diversity**

A qualitative study by Jean Anyon (1980) investigated how schoolwork compared amongst

students of different social classes and to see if there was a more hidden curriculum underlying the

schoolwork.  To collect data, she performed observations, interviews with fifth grade students,

teachers, principals, district administrators, and analyses of curriculum and materials.  Throughout

her research, she broke down the schools that were part of the study into categories based on the

socioeconomic status of the families that typically attend each school: "working-class," "middle-

class," "affluent professional," and "executive elite."  In these schools, she found there was a hidden

curriculum that was preparing the students to the type of work typical of the class they were raised

in (Anyon, 1980).  This shows that though many people talk about the social mobility of the

American Dream, schools have been aiding students in staying in the class they were raised in,

instead of aiding in upward mobility.

Oakes and Guiton (1995) conducted a qualitative research study to investigate how tracking

decisions are made in high schools.  They looked at case studies of four high schools with similar

state aide and standards, yet different compositions of students.  Researchers looked at handbooks,

course offerings, and enrollment procedures in addition to interviews and observations at each

school.  They interviewed students, faculty, and administrators.   They found that though faculty

and staff viewed that their tracking systems were fair and equitable for all students, they appeared to

be favoring White and Asian middle-class students.  Students who had equal grades seemed to be

tracked differently based on the SES and racial background.  They discovered that the issue was

much more complex.  Students who were Latino were taking lower level courses because they were

not being provided with the academic supports that they needed, and many African American

students and students with a low socioeconomic status were taking vocational courses because it

appeared to lead to more job stability.  Parents of white middle-class students would take advantage

of a waiver that would allow them to sign their children up for more advanced courses even if they

had not gotten the grades required for those courses.  Students also have the option to take classes

based on their interest, and many students opted out of high level courses even though they were

eligible for them (Oakes & Guiton, 1995).   This research shows how nuanced and complex the

issues around curriculum and tracking are in the schools.

In science classrooms, it is important to support the inquiry for all students, including our

female students.  In some many science classrooms when students are split into groups, the male

students are *doing* the science, while female students are filling in the data tables. In other cases,

female students are charged with traditionally caretaking roles in the classroom by aiding students

catch up who were absent. (Scantlebury, 2014)**.**  This pulls students away from the main classroom

instruction and opportunity to engage and explore scientific practices.

## Methodology

**Research Question**

In our investigation of beginning and intermediate science teachers we posed the following

overarching research question: What is the contribution from teaching experience, the teacher

preparation program, the classroom level, the gender of the students, and the diversity of the

students to the level of inquiry enacted in the classroom? Specifically, does the teacher preparation

program with higher levels of SMK make a difference to the level of inquiry-based science lessons?

Do middle and high school teachers implement inquiry-based lessons at the same level? And

finally, does the diversity of students affect the level of inquiry-based science lessons that teachers provide?

**Context**

There were 61 participants in this study, all graduates of a large, 4-year university in the Midwest's teacher education program. Of the 656 lessons, 209 were taught by graduates with a BA in secondary science education who met the minimum state requirements for either a single-subject or a general science teaching endorsement and 447 lessons were taught by graduates with a MA in education who all had at least an undergraduate degree in an area of science and were all certified in at least one single-subject (i.e., biology, chemistry). In the sample, 481 lessons were taught in middle schools and 175 lessons were taught in high schools. Teachers' experience ranged from their first-year teaching through their sixth year of teaching, with an average experience of 2.8 years in the classroom and standard deviation of 1.45.

The data for this study was gathered over the course of multiple academic years, 2015-16 through 2017-18. During the classroom observations, field notes were taken by a member of the research team, which were later coded using the EQUIP rubric. The observations were conducted by the PI and five graduate students. To ensure reliability specific points throughout the year, the researchers conducted a calibration. During this process, they observed videos of science lessons, discussed their coding, observed teachers in all possible pairs for multiple lessons. In these conversations the observers compared and discussed their results from the EQUIP rubric to come to consensus prior to making observations alone.

**Research Approach**

A quantitative approach to investing the teachers' use of inquiry-based instruction. A multiple regression was conducted on each of the EQUIP rubric items using the following factors:

(a) program (BA/MA), (b) classroom level (MS/HS), (c) teacher's years of experience, (d) gender of the students, and (e) diversity index of the classroom. Multiple regressions were also run on the aggregated assessment, instruction, curriculum, and discourse scales, in addition to an aggregated total of the entire EQUIP using the same factors listed above.

In the field notes for each observed science lesson, observers also counted the number of students, did their best to determine sex and race. Without further IRB approvals we were not able to determine SES for individual students and were only able to use publically-available school-level data, specifically free and reduced lunch participation as a proxy for SES. We entered these data as metadata for each science lesson in our spreadsheets for analysis.

## Results

### Descriptive Statistics of the Data

The table below gives the break down for the number of lessons for each teacher-level characteristic out of the 656 total lessons.

Table 1. *Number of lessons for each of the teacher-level characteristics for this analysis*

|  |  | # of lessons |
|---|---|---|
| Level of School | MS | 481 |
|  | HS | 175 |
| Program | BA | 209 |
|  | MA | 447 |
| Certification | In Field | 513 |
|  | Out of Field | 143 |

The table below shows the number of teachers observed each year in the study. There were 23% of teachers participated for one year, 36% of teachers participated for two years, and 41% of teachers participated for all three years.

Table 2. *Teachers observed by year of study*

| 2015-16 | 2016-17 | 2017-18 |
|---|---|---|
| **40 teachers** from both programs were recruited (2:1 MAst to undergrad program) **234 classroom observations** | **38 teachers** from both programs were recruited (2:1 MAst to undergrad program) **268 classroom observations** | **42 teachers** from both programs were recruited (2:1 MAst to undergrad program) **241 classroom observations** |

**Calculation of the Diversity Index**

To calculate and compare the levels of classroom diversity we used a generalized variance approach or the Absolute Diversity Index. In this method, diversity was calculated using the following equation: $GV = 1 - \sum_{i=1}^{C} P_i^2$, where $P_i$ values are the percentages of different demographic groups in the classrooms (Budescu & Budescu, 2012; Summers, Jackson, Woodward, Jones, & Dryer, 2011; Tam & Bassett, 2004). This system for measuring diversity has been commonly used in a range of fields including psychology and medicine.

**Regression Analyses**

A multiple regression was completed on the raw total scores of the EQUIP, the results showed that the predictors accounted for 10.5% of the variance ($R^2 = 0.105$, $F(5,649) = 15.18$, $p < 0.001$). The predictors teaching experience ($\beta = 0.230$, $p < 0.01$), teaching level ($\beta = -0.210$, $p < 0.001$), teacher preparation program ($\beta = 0.203$, $p < 0.001$), and student sex ($\beta = -0.122$, $p < 0.01$) all significantly contributed to the model. The diversity index of the students ($\beta = -0.041$, $p > 0.1$) did not contribute to the model.

A regression analysis was also completed on each of the regression analyses were found to be significant at a *p*-value of 0.000. On the assessment factors scale items A1, A3, and A5 were unable to be analyzed with multiple regression analyses due to the limited variability of the data.

This was also the case for one item on the instructional factors scale I5, and one item on the discourse factors scale D1. Specifically, these EQUIP rubric items referred to: prior knowledge (A1), student reflection (A3), the role of assessing (A5), knowledge acquisition (I5), and the questioning level employed (D1).

The diversity index of the classroom was only found to be significant with C4 in the EQUIP, which refers to the flexibility of how students record and organize information. In all other analyses of the EQUIP items, the diversity index of the classroom was found to be non-significant. This means that in most cases, the diversity index of the classroom had no impact on the level of inquiry the teachers in the study used in their classrooms.

The teacher preparation program was found to be significant in all items and aggregate totals, except for C4 (the flexibility of how students record and organize information). This means that teachers prepared through a MA degree enacted higher levels of inquiry in their classroom than teachers with a BA in secondary science education who only had minimal (i.e., a minor's worth) science content knowledge with respect to the remaining testable items in the EQUIP.

We found that the level of the classroom, middle school vs. high school to be significant with respect to most items on the EQUIP. It was only found to be non-significant for A2 and C4, which were items that concerned "conceptual development" and "organizing and recording information." In other words, inquiry was found more often in middle school classrooms than implemented by teachers teaching high school students in twelve items for the EQUIP.

The results for teaching experience was found to be significant on seven out of the fourteen testable EQUIP items. Out of the four possible discourse items, teaching experience was significant in all of them. When teaching experience was found to be significant, teachers with more

experience tended to enact more robust inquiry-based instruction in the classroom, especially with respect to incorporating more inquiry-based discourse.

We also investigated the potential effect of students' sex on the teachers' classroom instruction. Out of the fourteen testable EQUIP items, the sex of the students was significant for eight items. For these items, we found that classrooms with fewer female students had more inquiry-based strategies.

## Discussion

From the results of the analysis, the diversity of the students' race did not correlate to the use of inquiry-based instruction in the classroom. This may be related to the fact that the participants in this study were strategically placed in high needs locations throughout their field placements and student teaching experiences.

The results showed that teaching experience overall was found to be significant in the EQUIP total raw score, as well as for many of the individual EQUIP items, especially with respect to the discourse scales. This will require more investigation to see how teachers develop and evolve their use of inquiry with respect to discourse over time in the classroom.

The analysis showed that middle schoolers experienced more inquiry-based lessons than high schoolers. This will also require further investigation to make sure that these results are not be confounded with teaching in-field and out of field. The analysis will be run in the future with the middle school lessons broken down into the disciplines of physics, biology, chemistry, and earth science.

Sex was found to be a significant factor in the overall EQUIP raw score regression, as well as, for most of the EQUIP items. This result requires further analysis to see if this is due to student

sex, or if it was related more to the subject matter. If a subject such as physics was using more inquiry than biology, it would confound the results.

Program was found to be a significant factor for almost all of the EQUIP items. These results hold true with prior analysis (Lucas & Lewis, 2017). Participants with a master's in education also have a bachelors in their subject matter, such as physics, biology, or chemistry. Their subject matter knowledge gives them more flexibility in their ability to incorporate more reform-based teaching practices. The students who graduated with a bachelor's in science teaching have an equivalent of a minor in their chosen subject matter. As they do have the same depth of knowledge, they tend to incorporate less inquiry practices into the classroom.

### Conclusions

As with previous studies, this analysis showed the importance of teachers have a strong understanding of their content knowledge in order to be able to incorporate more inquiry-based instruction into their classroom.

For future research, we will be running analysis without the lessons that are out of field for the teachers based on their certification area to reduce confounds. We also will be running the analysis with the middle school lessons disaggregated into disciplines. By including the disciplines into the analysis, we aim to get a fuller picture about how student sex correlates with inquiry-based practices. We also aim to expand the analysis to see how diversity in the classroom relates to the implementation of science and engineering practices.

Some limitations of the research are that as this was a sample of convenience, the teachers were graduates of the same university and approximately 90% of them teach in the same state. This could impact the generalizability. The undergraduates of the teacher preparation program have not had any training of working with English language learners, which could add a confound. The

observations were targeted towards lessons that were taught in field, there was a lack of access to student level data per IRB so all student-level characteristics were estimated.

## Acknowledgements

**Paper #3: Validation of the Discourse in Inquiry Science Classrooms (DiISC) Instrument**

Lyrica Lucas, Brandon Helding, & Elizabeth Lewis

**Abstract**

Measuring inquiry-based science instruction is a relevant project to the science education reform movement as the new science education standards outline the practices that promote learning through an inquiry approach. In this paper, we propose that the *Discourse in Inquiry Science Classrooms* (DiISC) instrument can contribute to research in inquiry-based science instruction by addressing latent constructs on inquiry, oral discourse, writing, academic language development, and use of learning principles. Since the DiISC was developed within a specific professional development program, we present new evidence associated with aspects of a validity argument for the instrument. We used 660 coded science lessons to analyze the factor structure of the DiISC and to investigate its correlation with the *Electronic Quality of Inquiry Protocol* (EQUIP). A semi-structured interview of four raters of a video recorded lesson was used to collect and analyze data pertaining to substantive validity evidence.

**Introduction**

Observation and assessment instruments must be associated with strong validity and reliability arguments in order for researchers who use those measures to produce consistent, replicable, and generalizable results (AERA, APA, & NCME, 2014). This is particularly important as reform-based initiatives in science education emphasize the need for high quality, inquiry-based instruction and the research that undergirds it. As a consequence of using measures without adequate validity or reliability measures, researchers' efforts in studying classroom instruction are inevitably limited.

We attempted to address this need in our longitudinal study on beginning science teachers, using the *Electronic Quality of Inquiry Protocol* (EQUIP), an instrument with an established validity argument, to analyze the quality of teachers' inquiry-based classroom instruction (Marshall, Smart, & Horton, 2010). We are improving our investigation of inquiry-based science instruction by another assessment entitled the *Discourse in Inquiry Science Classrooms* (DiISC). We are using the DiISC to investigate a variety of instructional, inquiry-based factors that are not addressed with only the EQUIP measure.

**Literature Review**

Research instruments with high quality validity arguments are required for the high-quality

research needed to study inquiry-based instruction. In this effort, the DiISC was developed by

Baker's team (2008) for the *Communication in Science Inquiry Project* (CISIP) to measure

teachers' fostering science classroom discourse communities. (Lewis, Baker, Bueno Watts, and van

der Hoeven Kraft, 2016; Lewis, Baker, and Helding, 2015). The DiISC is a classroom observation

instrument designed to measure teachers' practices in creating a science classroom discourse

community through inquiry, oral and written discourse, academic learning strategies and learning

principles.  In this project we will use multiple measurement instruments, as well as present a

modern validity argument for the use of the DiISC instrument that draws on the already-validated

EQUIP instrument.  Each are described below.

**Electronic Quality of Inquiry Protocol (EQUIP) Instrument**

In order to better understand teacher behaviors associated with inquiry-based instruction, we

investigated the correlations between the DiISC and EQUIP. EQUIP measures the quality of inquiry

in an observed science lesson. The instrument consists of 19 items that measure inquiry-based

instruction based on four factors: (a) instructional strategies, (b) discourse, (c) assessment practices,

and (d) curricular features.

**The Discourse in Inquiry Science Classrooms (DiISC) Instrument**

The DiISC was initially developed at Arizona State University (ASU) and focuses on

observing teachers' instructional strategies that employ inquiry, discourse (oral and written),

academic language development, and learning principles (Baker, et al., 2008). Since the DiISC was

developed and validated within the context of a specific PD program, it is program-specific and

requires further scrutiny and development of an external validity argument for widespread use. This

work aims to establish elements of that validity argument to correspond with aspects of a larger, emerging validity argument (Kane, 1990).

**Inquiry in Classrooms and Measurement**

While both the EQUIP and DiISC instruments claim to measure constructs associated with inquiry-based instruction, these latent constructs are described by different observed variables. For example, DiISC items describe teachers' behavior and the instructional strategies they implement to support inquiry, oral discourse, writing, academic language development, and use of learning principles. Similarly, EQUIP items describe observed behaviors of teachers and students to assess the level of inquiry-based instruction in the areas of instruction, discourse, assessment, and curriculum. Unfortunately, the EQUIP instrument does not attend to issues of academic language development, related to equity issues for English language learners (Lee, 2017).

**Traditional View of Validity**

The classic notion of validity is it that is in the test itself. It is, therefore, part of how the test is used whenever it is used. This inherent validity is typically based on three validity issues (sometimes referred to in psychometric colloquialism as the 'holy trinity' of validity). It is based on content, construct, and criterion validity. This has traditionally drawn on content analyses, factor analyses, and correlations with extant criteria. In the last 30 years the very idea of validity has shifted toward it being an interpretive argument that is made of an instrument, associated with how it is used, and then adopted or not by other researchers when they choose to use that instrument in other studies and for other purposes (Kane, 1992; Messick, 1987, 1995).

**Unitary Concept of Validity**

The modern view of validity (or unitary concept of validity) reframes validity as an interpretive body of evidence, or preponderance of evidence, that suggests the appropriateness of

using a measurement instrument for a particular purpose. In particular, validity is no longer a

checklist, an inherent property of a measurement instrument, and is established through an ongoing

data collection and analysis process. The aspects of modern or unitary validity include content,

external, generalizability or predictability, structural, substantive, and consequential validity.

Messick (1994, pg. 9) describes them below.

> The content aspect of construct validity includes evidence of content relevance,
> representativeness, and technical quality. . . . The substantive aspect refers to
> theoretical rationales for the observed consistencies in test responses, including
> process models of task performance. . . , along with empirical evidence that the
> theoretical processes are actually engaged by respondents in the assessments tasks.
> The structural aspect appraises the fidelity of the scoring structure to the structure of
> the construct domain at issue. . . . The generalizability aspect examines the extent to
> which score properties and interpretations generalize to and across population groups,
> settings, and tasks . . . , including validity generalization of test-criterion relationships.
> . . . The external aspect includes convergent and discriminant evidence from multitrait-
> multimethod comparisons . . . , as well as evidence of criterion relevance and applied
> utility . . . The consequential aspect [included in a broader view of validity discussed
> below appraises the value implications of score interpretation as a basis for action as
> well as the actual and potential consequences of test use, especially in regard to sources
> of invalidity related to issues of bias, fairness, and distributive justice.

Specifically, in this paper with the DiISC, potentially enhances the way in which we are able

to capture inquiry-based, teacher practices, academic language development, and learning

principles. Multiple types of evidence are used and analyzed in varied ways to build, support, and

establish a strong validity argument (Kane, 1990), including recommendations for fairness and

fidelity in testing and interpreting results. Aspects of a validity argument include evidence for

content, external, generalizability, structural, substantive, and arguably consequential validity of an

assessment. Our purpose was to provide partial evidence for an emerging validity argument for the

DiISC instrument.

**Theoretical Overview**

The DiISC measures the creation of science classroom discourse communities (SCDC) to address communication in science and the language needs of all students. The SCDC model is one of situated learning (Lave & Wegner, 1992; Wegner, 1998) in which learning how to talk and write in science contributes to making meaning and the development of structured, coherent ideas (Kelly, 2007; Klein, 1999; Rivard & Straw, 2000 Yore, Hand & Prain, 1999). The DiISC also measures language principles and theories of Carrasquillo and Rodriquez (1996) and the Cognitive Academic Language Approach (Chamot & O'Malley, 1987) for academic language development.

The SCDC model is also based in social constructivism and emphasizes inquiry as a way to build knowledge (National Research Council, 1996). Within inquiry, there is an additional focus on the nature of scientific communication emphasizing text structures, and patterns of argumentation (Halliday & Martin, 1993). As teachers learn in the professional development about building a science classroom discourse community, there has also been a strong focus on learning for understanding science. Consequently, the DiISC was also designed to measure teachers' implementation of learning principles (Bransford, Brown, & Cocking, 2000; National Research Council, 2005). We present more of the key literature that has undergirded science education reforms in the last 20 years since the first national science education standards were published in 1996 (NRC, 1996).

**Scales on the DiISC Teacher Observation Instrument**

**Inquiry Scale**

Reform movements and the National Science Education Standards identify inquiry as essential to effective science teaching and student learning (National Research Council, 1996). In addition, the Inquiry scale addresses all of the practices of scientists and engineers describe in the Next Generation Science Standards (NGSS) Framework (i.e., asking questions, planning and

carrying out investigations, analyzing and interpreting data, constructing explanations, engaging in argument from evidence, developing models and using mathematics) (National Research council, 2012). Employing inquiry requires teachers to create an environment within which students engage in a set of complex cognitive processes (Windschitl, 2004). The Inquiry scale on the DiISC teacher observation instrument reflects the essential features of inquiry and measures the degree to which inquiry-based instruction takes place in a student-centered classroom where students explore the natural world with varying degrees of independence. The major consideration in developing items for this scale was to identify observable behaviors found in inquiry-oriented classrooms, including, but not limited to, aspects of the 5E instructional model (Bybee, 2009).

### Oral Discourse Scale

The Oral Discourse scale measures the degree to which teachers' instruction bridges students' everyday experiences and scientific discourse to create a science classroom discourse community. The scale focuses on whether the teacher is providing students with opportunities to build scientific vocabulary and engage in peer-to-peer discussions that support the construction of scientific arguments. It also focuses on whether the teacher is providing opportunities for students to explore the nature of scientific communication.

Scientific discourse in classrooms has been defined as knowing, doing, talking, reading, and writing (Moje, Collazo, Carillo, & Marx, 2001), or as the combination of scientific ways of talking, knowing, doing and using appropriate form of evidences (Lemke, 1990). Scientific discourse provides a vehicle for the social and cultural construction of knowledge (Alexopoulou & Driver, 1996; Kelly & Crawford, 1997; Kelly & Green, 1998; Kittleson & Southerland, 2004) through negotiation of meanings. Newton, Driver, and Osborne (1999) argue that in addition to conceptual understanding, discourse creates a scientific community in classrooms.

**Writing Scale**

The Writing scale measures the degree to which students have opportunities to pre-write, write, and share writing. These activities support acquiring the language patterns and vocabulary to communicate scientific ideas, use science notebooks, and the development of a science classroom discourse community. Several researchers assert that writing is both a reflection of conceptual understanding and a tool to generate understanding (Halliday & Martin, 1993; Lemke, 1990). In his review of the research, Rivard (1994) wrote that "students using appropriate writing-to-learn strategies are more aware of language usage, demonstrate better understanding and better recall, and show more complex thinking content" (p.975). Furthermore, explicit teaching of scientific writing helps students to organize relationships among factual information (Callaghan, Knapp, & Noble, 1999).

**Academic Language Development Scale**

The Academic Language Development scale measures the degree to which teaching supports scientific language development through the use of visual aids, supplemental resource materials, and clear instruction. It also measures the degree to which lessons build on students' everyday language and culture and provide opportunities for students to acquire scientific vocabulary. The specific items on the Academic Language Development scale reflect strategies adapted from Herell and Jorden (2003) as well as the research in science education that has addressed linguistically diverse students (Fradd & Lee, 1999; Lee & Fradd, 1996).

**Learning Principles Scale**

The Learning Principles scale measures the degree to which teaching provides opportunities for students to assess prior knowledge, make conceptual connections, and engage in metacognition. The scale also measures whether the teacher models scientific thinking, establishes community

norms, and promotes an academic focus that supports learning science. The Learning Principles

scale is the largest scale of the DiISC and is based upon the cognitive principles outlined in How

People Learn and How Students Learn (Bransford, Brown, & Cocking, 2000; National Research

Council, 2005). There is an emphasis on metacognition, as part of self-regulated learning, because

students must "develop the ability to take control of their own learning, consciously define learning

goals, and monitor their progress in achieving them" (National Research Council, 2005, p.4-10), as

well as the essential role of factual knowledge and conceptual frameworks in developing an

understanding of science (National Research Council, 2005). Items also measure timely and specific

feedback as a key element of formative assessment (Black & Williams, 1998) to guide students to

develop understanding.

**Methodology**

In this study, we examined several sources of evidence and conducted several analyses.

Although the DiISC was developed within the CISIP project, the items were first constructed from a

review of educational research to reflect alignment with science education standards promoting

social constructivism and science as inquiry (Baker, Lewis, Purzer, Watts, Uysal, Wong, Beard, &

Lang, 2009). In the users' manual they included a table of specifications and a description of the

domains of the instrument. This adequately supports the content validity of the DiISC.  This

analysis will further examine the external, structural, generalizability, substantive, and make

recommendations for the consequential validity of the DiISC.  Each is discussed in separate sections

below.

Before generating any validity evidence anew, we examined previously documented

development processes, reliability measures, and pilot results by the developers in a technical

report. They described an iterative design and evaluation process to establish the content, face,

construct, and concurrent validity of the DiISC. The developers reported a high intraclass correlation, $r$=0.90, to indicate strong agreement among raters. They also conducted an EFA using 204 classroom observations of middle and high school science teachers in their study. The EFA identified five factors, accounting for 46.1% of the total variance. Unfortunately, the DiISC lacked detailed validation evidence required for widespread use. As part of our collection of validity evidence we conducted several analyses. We add to this previous work with the aforementioned aspects of validity argument.

## Results and Conclusions

### Content Validity Argument

The test developers provided evidence of criterion validity using the correlation between classroom observation DiISC scores and *My Science Classroom Survey* given to 187 students. We use this text to support our content validity argument. Specifically, the survey is a measure of students' perceptions of their teachers' use of instructional strategies. The test developers found a statistically significant correlation between the DiISC scores and students' perceptions *($r$=0.80, $p$<0.01)*. As part of the validation process, we used a number of measures to investigate the correlations between the DiISC and EQUIP. Similarly, as already stated, previous work on the DiISC, in the users' manual they included a table of specifications and a description of the domains of the instrument. This adequately supports the content validity argument for the DiISC.

### External Validity Argument

The external validity argument in this case was based on the factor structures established for the both the EQUIP and DiISC instruments. The factor structure of each measure was used to generate factor scores for the raw scores on each instrument. Because the EQUIP already has an entant validity argument, its correlation with the DiISC was used as the external validity argument

for it.  This is a typical way of establishing external validity  and evidence for it (Kane, 1990, 2013).

In the parlance of a correlation in external validity argumentation the proportion of variance

explained indicates the true scores without systematic error (given naturally occurring error

associated with the multivariate regression).

In this case, the DiISC factor scores were used to predict the EQUIP factor scores.  In each

of the three, extracted DiISC factors (inquiry, discourse, learning principles), there was a

statistically significant correlation with the two EQUIP factors (instructional and curricular factors,

discourse and assessment factor) and the DiISC factors (inquiry, *Pillai's Trace*=0.63(2,652),

*p*<0.01; discourse, *Pillai's Trace*=0.04(2,652), *p*<0.01; learning principles, *Pillai's*

*Trace*=0.23(2,652), *p*<0.01).  Validity coefficients for each factor, respectively, are 0.63, 0.04, and

0.23.  This provides evidence for the external validity argument for the DiISC instrument.  This was

further supported by unpatterned standard errors in residual plots.

**Predictive or Generalizability Validity Argument**

The weakest part of the validity argument in this particular presentation is the predictive or

generalizability validity argument.  That said, the previous paper in this presentation has indicated

that analyses were conducted over several subgroups of the population or sampling in this case, in

question, and that issues of differential item function (DIF) were not prevalent.  Furthermore, cross

validation with the EQUIP predicted a clear regression line, but shrinkage or decay over time was

not yet analyzed.  This type of longitudinal analysis is ongoing, part of the larger research project,

and an element of the validity argument for which we have preliminary evidence but yet compelling

cases and preponderance of indicators to make strong conclusions.

Also, it should be considered that this validation process is part of a larger project that does

use a representative sample of BA teachers and MA teachers.  To this extent the results on the

EQUIP and DiISC are useful in making generalizations and have been done so in the larger SEM. This would indicate, via indices of model fit as well as minimal estimates of measurement error (in the SEM) that the factor scores associated with the DiISC are stable and potentially predictive or generalizable to a larger population of inservice teachers and career changers (in the MAst program).

**Structural Validity Argument**

A total of 660 sets of DiISC observation were used to conduct an Exploratory Factor Analysis (EFA). We used extracted factors using principal axis factoring and allowed factors to correlate with each other (using a PROMAX rotation). Preliminary results indicated that a three-factor solution was appropriate (Lewis et. al, under review). As a result, we extracted three factors from the DiISC observations and used those three factors in the external validity argument (above) to generate factor scores that were in turn correlated with the EQUIP factors scores. This extraction was supported by a scree plot. These three factors met the interpretability criterion.

Items and the fully rotated pattern matrix are shown in Table 1, showing the simple structure. The proportion of each variables variance accounted for by the factor analysis is shown in the communality column. We identified items with factor loadings > 0.30 on their respective factors. The overall Kaiser-Meyer-Olkin was 0.81 demonstrating that the sampling is adequate. Bartlett's test of sphericity was statistically significant ($p<0.001$) indicating that the data was factorizable. The primary loadings of most of the items in the original *inquiry* scale are in column 1. The primary loadings of most of the *oral discourse* items are in column 2 and the items about *writing and learning principles* are in column 3. Factorial validity of the three-scale solution appears to be supported by the simple structure of the matrix. Eleven items had loadings <0.30 and did not clearly load to any of the factors.

Table 1. *Factor Loadings for Items*

| DiISC Items | Factor | | | $h^2$ |
|---|---|---|---|---|
| | 1 | 2 | 3 | |
| 4. Opportunities for early stages of scientific exploration: making observations, recording data, and constructing logical representations (e.g., graphs) | 0.86 | 0.20 | 0.22 | .075 |
| 1. Teacher creates an environment that supports inquiry | 0.82 | 0.36 | 0.13 | 0.71 |
| 2. Teacher engages students in asking scientific questions for the purpose of investigation (hands-on or other means) | 0.71 | 0.10 | 0.15 | 0.54 |
| 5. Opportunities for later stages of scientific exploration: explaining phenomena via claims and evidence, making predictions, and/or building models | 0.68 | 0.38 | 0.31 | 0.50 |
| 28.  Teacher and/or students situate factual knowledge (experiences, ideas, data, and explanations to past lessons and/or real-world experiences) within a conceptual framework (fact to concept relationship) | 0.54 | 0.24 | 0.50 | 0.40 |
| 8. Teacher promotes peer-to-peer discussion | 0.53 | 0.19 | 0.38 | 0.32 |
| 24. Teacher provides instruction for interactions among students | 0.52 | 0.13 | 0.18 | 0.28 |
| 3. Opportunities for students to design and plan exploration of the natural world individually or in groups | 0.50 | 0.41 | 0.14 | 0.31 |
| 36. Teacher uses feedback strategies that have an academic focus (NOT just praise; "be more specific") | 0.42 | 0.31 | 0.40 | 0.27 |
| 6. Generating scientific arguments and constructing critical discourse about limits and sources of error | 0.28 | 0.25 | 0.04 | 0.12 |
| 20. Using visual aids and gestures to communicate with students | 0.26 | 0.06 | 0.24 | 0.09 |
| 29. Teacher provides opportunities for students to review key concepts (focus on the review, not the discourse) | -0.13 | 0.08 | 0.06 | 0.05 |
| 11. Teacher engages students in discussion that emphasizes the nature of science | 0.13 | 0.04 | 0.11 | 0.02 |

| | | | | |
|---|---|---|---|---|
| 12. Formal writing in a genre that reflects the nature of science | 0.17 | 0.74 | 0.05 | 0.58 |
| 13. Engaging students in prewriting associated with science concepts | 0.05 | 0.54 | 0.04 | 0.32 |
| 14. Engaging students in recursive writing processes using rubrics to review and revise | 0.17 | 0.52 | 0.05 | 0.28 |
| 15. Engaging students in writing to acquire the language patterns and vocabulary to communicate scientific ideas | 0.21 | 0.51 | 0.25 | 0.27 |
| 16. Teacher provides direct instruction in writing content, forms, and processes | 0.16 | 0.51 | 0.15 | 0.25 |
| 35. Communicating lesson expectations with guidelines (oral or written) , or rubrics, or exemplars | 0.30 | 0.44 | 0.23 | 0.22 |
| 32. Teacher provides students opportunities to develop awareness of their own learning strengths and challenges | 0.18 | 0.40 | 0.14 | 0.16 |
| 22. Teacher addresses multiple levels of academic language proficiency (differentiated instruction and/or assessment) | 0.08 | 0.33 | 0.08 | 0.11 |
| 33. Promoting executive control of learning (student choice about what and how they learn) | 0.21 | 0.29 | 0.13 | 0.10 |
| 17.  Engaging students in using science notebooks as a learning tool | 0.18 | 0.21 | 0.14 | 0.06 |
| 25. Uses supplemental resource material (Note: lesson could be done without these) | -0.01 | 0.05 | 0.04 | 0.01 |
| 18. Providing students opportunities to acquire vocabulary | 0.12 | 0.30 | 0.55 | 0.34 |
| 10. Teacher models scientific discourse and vocabulary | 0.07 | 0.01 | 0.54 | 0.32 |
| 30. Teaching with embedded metacognition for students to elaborate and summarize their understandings | 0.38 | 0.24 | 0.50 | 0.30 |
| 7. Teacher promotes discourse through questioning | 0.21 | 0.11 | 0.47 | 0.23 |

| | | | | |
|---|---|---|---|---|
| 19. Teacher uses clear instruction throughout lesson by modeling expectations | 0.36 | 0.28 | 0.44 | 0.25 |
| 9. Teacher (or instruction) bridges everyday experiences and scientific discourse | 0.04 | 0.07 | 0.38 | 0.16 |
| 31. Teaching self-monitoring for understanding (focus on direct instruction of strategies) | 0.24 | 0.20 | 0.36 | 0.15 |
| 23. Provides direct instruction for using academic learning strategies | 0.06 | 0.03 | 0.32 | 0.11 |
| 21. Building lesson on students' language (vernacular or non-English) OR culture | -0.02 | 0.12 | 0.23 | 0.07 |
| 26.  Accessing students' prior knowledge | 0.08 | -0.04 | 0.20 | 0.05 |
| 27. Teacher modifies instruction based on students' prior knowledge | 0.03 | 0.002 | 0.14 | 0.02 |
| 34. Teacher establishes or reminds students of community norms for discourse | -0.04 | 0.03 | -0.06 | 0.01 |

**Substantive Validity Argument**

To generate evidence of a substantive validity argument the same video recorded lesson was shown to four raters; that is, the four raters in this project were selected and interviewed.  They were given a semi-structured interview that focused on areas of transitions and ambiguity in the lesson. For example, when the teacher in the lesson switched from a video he had selected to group work, and then evaluation of the group work and back to the video, at each transition point the interviewees were asked about how they would code those instances using the DiISC.  It should be noted that the DiISC instrument is meant to encapsulate an entire lesson, rather than single elements of it, and accordingly each rater was asked to watch the entire video and code it using the EQUIP and DiISC instruments after the interview.

The interviews, as stated, were semi-structured with a focus on salient issues of the video recorded observations.  Additionally, these interviews included mini-tour and grand-tour questions

about the thinking associated with assigning codes to the lesson (Kvale, 2009). The interviews

involved think-aloud questions and adhered to think-aloud protocols (Ericsson & Simon, 1998).

The interviews were coded using categories and a priori themes associated with the instruments

(Miles, Huberman, & Huberman, 1994). They fit the categories well (provided in additional detail

in a longer paper). The scores on the overall video provided by the raters were highly correlated

using a linear regression. Unfortunately, a polychoric correlation that accommodated the ordinal

nature of the data was not sufficiently powered.

Nevertheless, there was strong evidence for the substantive validity argument for the DiISC

measure for those that were trained to use it. Of note, is that one rater that helped design the

original instrument, provided a great deal more information and insights into their coding compared

with the other raters. Even with the additional detail, however, it only confirmed the comments and

insights provided by the other three raters. This indicated that as raters improved in their

knowledge of the instrument, their ratings did not fundamentally change. More details are provided

in a larger document about the validation process.

## Discussion and Implications for Future Research

We were able to establish a strong body of evidence for the validity of the DiISC instrument

across standard aspects of a modern unitary validity argument: validity argument include evidence

for content, external, generalizability, structural, substantive, and arguably consequential validity of

an assessment. The content aspect of validity was completed by the DiISC developers, and the

structural and external validity were well established with factor analyses and correlations with the

EQUIP instrument. The substantive validity argument was well established through think-aloud

and semi-structured interviews with the raters in question. Further interviews will be necessary, of

course, with a larger body of raters when used in other contexts when more individuals are involved

in data collection, especially individuals without specialized knowledge in education, classrooms, and educational research. The generalizability or predictive validity is currently the weakest area of the overall validity argument but is bolstered by preliminary analyses and initial findings. This is an area for further research in the ongoing process of establishing and providing evidence for the valid use of the DiISC instrument in widespread educational research.

It should be noted that throughout this investigation, we have neglected consequential validity. We do so on philosophical grounds. While we can train individuals to use the DiISC and even monitor how they are using it, the potential for misuse is pervasive and constant. In this way, until the DiISC instrument grows in popularity, is clearly misused or used to mislead, we do not take responsibility for how others will interpret our validity argument. We feel our argument is persuasive and compelling, and that it should guide the prudent use of the DiISC instrument as part of high-quality educational research that is badly needed.

**Paper #4: Modelling Beginning Science Teachers' Inquiry-based Science Teaching**

Elizabeth Lewis, Brandon Helding, Lyrica Lucas, Amy Tankersley, & Elizabeth Hasseler

## Introduction

With new national science education standards (NGSS Lead States, 2013), it is critical to understand how to educate and support science teachers who are capable of advancing science education reform priorities. The *Next General Science Standards* (NGSS) require science teachers to be fluid in their selection, development, and implementation of curriculum within three dimensions of science learning.  These dimensions include: (a) disciplinary core ideas, (b) scientific and engineering practices, and (c) cross-cutting concepts.  The *NGSS* three dimensions of science learning articulate aspects of science content knowledge, scientific methods, and the nature of science in an integrated approach to learning science.  All dimensions require that science teachers have a strong understanding of science themselves and effective ways of teaching that are grounded in learning theory.  In the first national science education standards (NSES) the way to enact reformed-based science teaching was referred to as inquiry-based instruction.  Models of inquiry-based instruction have been around for decades, but have been difficult to achieve in practice (Cuban, 1992).  To help meet our goals for scientific literacy the NGSS learning objectives are more explicit and fine-grained than the NSES (NRC, 1996) and leverage three-dimensional learning as a means for achieving success.

This study was specifically designed to investigate the inquiry-based teaching practices of beginning science teachers with a range of in-field content knowledge and the relationship to exemplary, reform-based instruction using multivariate growth Structural Equations Modeling (SEM) based upon a 3-year longitudinal dataset with 660 classroom observations.

## Relevant Literature

Bianchini (2012) found that little is known about the science teaching induction period, and recommended that researchers produce more studies that follow beginning science teachers from their preservice teacher education preparation into the classroom; specifically focusing on their beginning instructional practices, and trace connections, or lack thereof, over induction training, and student learning. Our research contributes to understanding how to design effective science teacher preparation programs (TPPs) that result in teachers who can address the long-standing goal through national science education standards of scientifically literate citizens, as well as targeting areas of need for professional development.

When we consider what resources and efforts that are necessary to meet the vision of the NGSS and goal of robust scientific literacy for all students, it will require many things, but one of the most crucial elements is for students to have access to well-prepared science teachers. Unfortunately, the problem of out-of-field teaching has been a significant issue in many states that has undermined the capacity of teachers to be able to deliver robust science education, even at the secondary level.  Out-of-field teaching occurs when science teachers who are certified in one core area (e.g., biology) are assigned to teach a different science (e.g., Earth and space science). A recent study by Nixon, et al (2017) showed that only about one third of science teachers in their first five years are assigned to teach in-field by their administrators. They also reported that about 20% of teaching assignments were entirely out-of-field and about 43% of assignments were some

combination of in-field and out-of-field. When teachers teach out-of-field they lack confidence and subject matter knowledge that is necessary to teach using inquiry-based approaches (Treagust, 2014) and they are less likely to recognize student misconceptions and more likely to teach oversimplified content (Sadler & Sonnert, 2016; Hashweh, 1987).

Some of our recent work has focused on science teacher preparation, certification, and teachers' misconceptions about core physical science concepts, specifically chemistry and physics (Lewis, Rivero, Musson, Lucas, and Helding, 2019). For example, we found that teachers needed at least 30 college-level credit hours in chemistry at a 3.2 GPA in order to reliably pass a test of common high school level misconceptions. Furthermore, we were able to connect teachers' subject matter knowledge (SMK) to the level of inquiry-based teaching used in their classrooms; in predicting inquiry-based teaching practices the total number of chemistry credit hours taken by a teacher accounted for 19% of the variance in their use of inquiry in their science lessons (*partial $\eta^2$ = 0.190*) (Lewis, et al, 2018). Thus, while robust content knowledge is at the root of successful and effective teaching, it clearly is not the only important variable that needs to be investigated.

**Conceptual Framework**

In addition to the overall conceptual framework we used for the entire related paperset, we also considered other frameworks. In particular, to assist our grouping and consideration of the many teaching-related variables that have been identified in other education research, we used Cuban's (1992) framework of internal and external variables that influence curriculum change in American schools. This allowed us to identify exogenous and endogenous variables when specifying the SEM. We also embraced a constructivist stance toward learning science as the goal for inquiry-based instruction, having been adopted and explicated in national standards for over 20 years (Bybee, 2011).

## Research Approach and Methods

We used an exploratory approach to data analysis to investigate beginning science teachers' reform-based teaching practices, specifically building an exploratory SEM. Follow-up tests were conducted and are described below.

**Primary Research Questions**

We focused on beginning science teachers' teaching practices through the following research questions:

1. To what degree are teachers' practices reform-based (i.e., inquiry-based)?
   a. Do science teachers' inquiry-based instruction change over time?
   b. And if so, what are the significant variables that contribute to this change?
2. Is there a difference between lessons by teachers with less or more teaching experience?
3. Is there a difference between lessons that feature in-field (e.g., highly qualified teachers) and out-of-field certified teachers?
4. Do middle or high school teachers enact greater inquiry-based instruction?

**Context**

   We recruited secondary science teacher program graduates (*n*= 62) from a U.S. Midwestern 4-year state university. Teacher participants graduated from either a 4-year undergraduate program (*n*=25) or a 14-month master's degree in education with initial science teacher certification program (MAT) (*n*=37). Unlike the undergraduate program, the MAT program recruited teacher candidates who had earned at least a bachelor's degree in a scientific field and combined coursework required for teacher certification, graduate-level courses (with a capstone research project), and extensive (650+ hours) clinical experiences. Teacher graduates from the undergraduate program had no more than a minor in one area of science and may have pursued a general science, a so-called "broad field," endorsement.  Greater than one-half (58%) of graduates from both programs taught at high-needs schools, which represented 60.4% (29 of 48) of all schools in the sample.  High needs schools are defined as having more than 40 percent of students who qualified for Free or Reduced Lunch (FRL).

**Sample and Data**

   In this 3-year longitudinal study of TPP graduates, we observed up to six science lessons per teacher in each of the three academic years (2015-16 to 2017-18) and coded a total of *n*=660 science lessons by teachers with 0 to 6 years of teaching experience with a range of subject matter knowledge. We used the EQUIP instrument (Marshall, Horton, Smart, & Llewellyn, 2008) and the Discourse in Inquiry Science Classrooms (DiISC) instrument (Baker, et al., 2008).  In Paper #3 we provide a preliminary validity argument for the DiISC instrument, and therefore use that argument to justify its inclusion in the following SEM as a useful and accurate description of teachers' practices.  Additionally, we analyzed course transcripts, administered a teaching self-efficacy survey (Tschannen-Moran & Hoy, 2001), and collected responses to a questionnaire about teachers' ongoing PD activities after they left their teacher education program.

**Model-building Method**

   To analyze teachers' instructional patterns over time we used an SEM; we built a multivariate growth SEM with a combination with data available across the years of data collection. The model was built in and estimated in SPSS AMOS software. The complete specification picture is too large for this length of document. In an effort to explain how the model was specified, we have nonetheless included a simplified version with (co)variances between variables (exogenous and endogenous) overlapping in physical space when possible. The model is shown in Figure 2.  Of note is that variables that were either deemed insignificant theoretically, determined to be statistically insignificant, or superfluous were omitted from the complete model.  That said, the estimate of effects associated with statistically significant and meaningful predictors of teachers' reform-based science practices, and their change over time did include those factors.

   The full list of variables included in the model are shown below in Table 1.  Follow-up MANOVAs were conducted and focused on significant variables that predicted teacher inquiry-based instruction.  Specifically, we examined teaching experience, school level or type (e.g., middle or high school), in- and out-of-field teaching (e.g., a biology teacher teaching biology or physics, respectively), and potential interactions between school level and in- and out-of-field teaching.

Table 1. *Variables Included in Model*

| Variable | Description | Variable | Description |
|---|---|---|---|
| M0 | Teaching Self-efficacy survey 2015 | D11 | DiISC 2015-2016: Inquiry |
| M1 | Teaching Self-efficacy survey 2016 | D12 | DiISC 2015-2016: Discourse |
| M2 | Teaching Self-efficacy survey 2017 | D13 | DiISC 2015-2016: Learning Princ. |
| M3 | Teaching Self-efficacy survey 2018 | D21 | DiISC 2016-2017: Inquiry |
| C1DA | 2015-2016: EQUIP: Discourse & Assessment | D22 | DiISC 2016-2017: Discourse |
| C1CC | 2015-2016: EQUIP: Curricular Choices | D23 | DiISC 2016-2017: Learning Princ. |
| C2DA | 2016-2017: EQUIP: Discourse & Assessment | D31 | DiISC 2017-2018: Inquiry |
| C2CC | 2016-2017: EQUIP: Curricular Choices | D32 | DiISC 2017-2018: Discourse |
| C3DA | 2017-2018: EQUIP: Discourse & Assessment | D33 | DiISC 2017-2018: Learning Princ. |
| C3CC | 2017-2018: EQUIP: Curricular Choices | TSP | Teacher Preparation Program |
| T1FRL | Free Reduced Lunch rate: 2015-2016 | TSEH | Education Credit Hours |
| T1EXP | Teacher Experience: 2015-2016 | T2EXP | Teacher Experience: 2016-2017 |
| T1DEV | 2015-2016 Professional Development | T2DEV | 2016-2017 Professional Development |
| T2FRL | Free Reduced Lunch rate: 2016-2017 | T3FRL | Free Reduced Lunch rate: 2017-2018 |
| TSCE | Teacher Certification Type | T3EXP | Teacher Experience: 2017-2018 |
| | | T3DEV | 2017-2018 Professional Development |

## Results

The results are organized in three sections below. First, the overall SEM is presented with general findings, specific interpretations, and important areas for follow-up research. Those areas will be addressed with follow-up MANOVAs. There will be other areas for even further investigation, which are discussed with conclusions after the results.

### Overall SEM Results

The model was clearly specified using the governing research questions of this research project and the intent to identify the predictors of reformed-based teacher practices, if those practices change over time, and what predicted those changes over time. We specified and estimated our model in SPSS AMOS. The model fit well and with statistical significance (*CMIN* = 1593.76, d$f$ = 496, $p < 0.01$; $X^2$ (346) = 730.94, $p < 0.01$). Importantly, it even fit well when accounting for model complexity (*RMSEA* = 0.092, 90 CI = 0.087, 0.097). While this parsimony adjusted misfit is slightly higher than recommendations by Brown and Cudeck (1993), it nevertheless passed an exact test of model fit (*PCLOSE* < 0.01). Also, we found a statistically significant reductions in AIC and BCC misfit. The overall interpretation of all model fit indices indicated a preponderance of evidence that the model fit well (the entire model is shown in Figure 2. A simplified version is shown in Figure 3.

**Interpreting Specific Standardized Estimates (Estimated Coefficients in SEM)**

The estimated, statistically significant coefficients are provided in Table 2, arranged from largest to smallest. In the model specification it should be noted that the path between the intercept and each latent or measured predictor was restricted to 1 in all cases; similarly, variables paths that corresponded with 2015-2016 were restricted to 1, 2016-2017 to 2, and 2017-2018 to 3. All other paths, errors terms, and disturbance terms were allowed to vary freely, but were not correlated with one another to avoid biased estimates of model overfit.

Table 2. *Estimated, Statistically Significant Regression Coefficients*

| Factor 1 | Factor 2 | Estimate | *p*-value |
|---|---|---|---|
| 2015-2016: EQUIP: Curricular Choices | Teacher Program | 0.959 | >0.001 |
| Teaching Self-efficacy survey 2017 | Teacher Program | 0.557 | 0.013 |
| 2015-2016: EQUIP: Discourse & Assessment | Teacher Program | 0.545 | 0.002 |
| DiISC 2017-2018: Learning Principles | 2016-2017 Prof. Develop. | 0.41 | >0.001 |
| DiISC 2017-2018: Discourse | 2016-2017 Prof. Develop. | 0.278 | 0.007 |
| 2016-2017: EQUIP: Curricular Choices | 2016-2017 Prof. Develop. | 0.241 | 0.01 |
| Teaching Self-efficacy survey 2017 | Teacher Exp: 2015-2016 | 0.219 | 0.041 |
| DiISC 2017-2018: Inquiry | 2016-2017 Prof. Develop. | 0.211 | 0.044 |
| DiISC 2015-2016: Inquiry | Education Credit Hours | 0.069 | 0.008 |
| DiISC 2015-2016: Learning Principles | FRL rate: 2015-2016 | -0.009 | 0.013 |
| DiISC 2015-2016: Discourse | FRL rate: 2015-2016 | -0.017 | >0.001 |
| DiISC 2016-2017: Learning Principles | Education Credit Hours | -0.054 | 0.035 |
| 2015-2016: EQUIP: Discourse & Assessment | Education Credit Hours | -0.068 | 0.012 |
| 2015-2016: EQUIP: Curricular Choices | Education Credit Hours | -0.073 | 0.004 |
| DiISC 2016-2017: Inquiry | Education Credit Hours | -0.131 | >0.001 |
| DiISC 2015-2016: Learning Principles | 2015-2016 Prof. Develop. | -0.219 | 0.044 |
| Teaching Self-efficacy survey 2018 | 2016-2017 Prof. Develop. | -0.262 | 0.017 |
| DiISC 2016-2017: Inquiry | Teacher Certification Type | -0.358 | 0.018 |
| 2016-2017: EQUIP: Curricular Choices | Teacher Program | -0.432 | 0.009 |
| 2015-2016: EQUIP: Discourse & Assessment | Teacher Certification Type | -0.466 | 0.005 |
| 2016-2017: EQUIP: Discourse & Assessment | Teacher Program | -0.469 | 0.006 |
| 2015-2016: EQUIP: Curricular Choices | Teacher Certification Type | -0.587 | >0.001 |
| DiISC 2015-2016: Inquiry | Teacher Certification Type | -0.706 | >0.001 |

In Table 2 there are several interesting results. First, the inquiry-based practices of teachers remained similar over the course of the study (with a change over time of 0.01) and was initially, highly correlated with either Teacher Program (MAT) or later on-going professional development.

Specifically, the EQUIP 2015-2016 measure of curricular choices was nearly perfectly correlated with a teachers' program, meaning that teachers from the MAT program used higher levels of inquiry-based instruction. Second, in the SEM, teacher program was important in not only the strongest three correlations (on the paths between teacher program and EQUIP curricular choices subscale, self-efficacy survey, and EQUIP discourse and assessment subscale), but was also associated with increased inquiry-based instruction when combined with professional development over time.

Third, having membership in a high quality teacher preparation program (i.e., MAT program) coupled with ongoing professional development was important for inquiry-based instruction once teachers had been in classrooms longer. Specifically, the amount of ongoing professional development that teachers received was positively correlated with inquiry-based instruction on the DiISC learning principles subscale, discourse subscale, and EQUIP curricular choices subscale in the 2017-2018 not the 2016-2017 school year. This has clear implications for the overall preparation of teachers and guided follow-up analyses.

This provided us a nuanced insight into teachers' inquiry-based instruction. While both teacher preparation program and ongoing professional development were important, they were important differently over time. They both occurred, though, and as a result they are inseparable as part of the education that any teacher received when they were in the MAT program (which was associated with greater inquiry-based instruction). As a result, we argue that teachers need *both* a high-quality teacher preparation program with robust subject matter knowledge as well as ongoing professional development. It also has several implications for follow-up analyses. They are described in detail below.

**Follow-up Tests**

Follow-up tests were conducted to examine the specific factors that predicted inquiry-based instruction in teachers' instruction using targeted MANOVAs. We investigated amount of teaching experience recoded in two categories: 0-3 years or 4-6 years of experience and teacher certification area. Accordingly, we continued with teaching assignments (in- or out-of-field teaching placement), the level the school in which teachers taught (middle vs. high school). The results of each follow-up investigation is provided below. The distribution of science lessons observed that were in either middle or high schools or in- and out-of-field teaching assignments is provided below in Table 3.

Table 3. *Distribution of Observed Science Lessons Used in the Study*

| | | Observed Science Lessons | | | |
|---|---|---|---|---|---|
| | | **2015-16** ($n$=212) | **2016-17** ($n$=220) | **2017-18** ($n$=228) | **Total** ($n$=660) |
| Certification and teaching assignment | In-field, $n$ (%) | 174 (82.1) | 153 (69.6) | 186 (81.6) | 513 (77.7) |
| | Out-of-field, $n$ (%) | 38 (17.9) | 67 (30.5) | 42 (18.4) | 147 (22.3) |
| Lesson level | MS, $n$ (%) | 53 (25.0) | 55 (25.0) | 66 (29.0) | 174 (23.4) |
| | HS, $n$ (%) | 159 (75.0) | 165 (75.0) | 162 (71.1) | 486 (73.6) |

The follow-up analyses yielded three primary results per the research questions previously noted. First, there was a statistically significant relationship between inquiry-based instruction and teacher experience (*Pillai's Trace* (5,651) = 0.37, *p*<0.01), indicating that more experienced teachers used more inquiry in their lessons. Second, there was a statistically significant relationship between inquiry-based lessons and in-field and out-of-field teachers for those lessons (*Pillai's Trace* (5,651) = 0.49, *p*<0.01) in which in-field single-subject science teachers delivered lessons featuring greater inquiry-based instruction. Third, there was a statistically significant relationship between inquiry-based lessons and middle and high school lessons, (*Pillai's Trace* (15,1953) = 0.38, *p*<0.01) in which high school teachers enacted lessons using greater levels of inquiry. The predictors were then entered into a common, omnibus MANOVA.

In the omnibus MANOVA, each of the aforementioned predictors statistically significantly predicted inquiry-based instruction with varying levels of unique variance account for, with one interaction term:

1. teaching experience: *Pillai's Trace* (5,646) =0 .06, *p*<0.01, *partial $\eta^2$* = 0.063;
2. school level: *Pillai's Trace*(15,1944) = 0.26, *p*<0.01, *partial $\eta^2$* = 0.086)
3. in- and out-of-field teaching: *Pillai's Trace*(5,646) = 0.32, *p*<0.01, *partial $\eta^2$* = 0.32);
4. and an interaction between school level and in-/out-field: *Pillai's Trace* (5,646) = 0.17, *p*<0.01, *partial $\eta^2$* = 0.17.

These results are discussed below.

## Discussion

The SEM model in Figure 1 was estimated and in Table 2 only the statistically significant path coefficients were included. This is not to mean that the many other paths in the model were not important, in fact they are vital to the overall model interpretation and possible future respecifications. Specific to our research questions, the MAT teachers' practices were highly correlated with inquiry-based instructional practices. Similarly, the program participation improved those teachers reported self-efficacy and ongoing professional development helped teachers persist in their use of inquiry-based instruction. Follow-up analyses indicated that teacher experience, school type (middle vs. high school) and in- and out-of-field teaching predict inquiry-based instruction as well.

Interestingly, these were not found as statistically significant in the larger SEM. The implication is that respecifications of the SEM needs to address not only the statistically significant elements of the SEM, seeking a parsimonious and interpretable outcome, but also that the SEM was indeed exploratory. Future, confirmative SEMs will include not only those factors previously identified in the SEM, but also the factors identified in the follow-up MANOVAs (e.g., experience, school level, etc.). In fact, respecifications based on these theoretical and practical lessons will help us build an overall SEM that profiles teachers as they use inquiry-based instruction in the context of various teacher preparation programs, types of teaching assignments, and their professional development opportunities once they are in-service teachers.

In other words, it may be surprising to see that educational coursework was either only a minimal contributor, or not a significant factor (Table 2). Because we had targeted the major difference in the two routes of teacher preparation as the amount of science coursework teachers had taken, there was less difference between the educational coursework than the science

coursework. Had we recruited for example, a third group of teachers who were emergency certified with little or no education coursework we might have found something different. Thus, this is a limitation of the study as it does not address a broad enough range of education coursework to determine what might be a practical minimum amount that is necessary for effective science teaching.

**Conclusions and Implications for Future Research**

The primary conclusion is that both high quality teacher preparation and ongoing professional development are critical to inquiry-based instruction. Importantly, however, they are best when combined together, and intertwined in the same models of teacher learning. This is limited by several elements of the analysis, but also will lead us in several interesting directions for future research. These future, targeted investigations will be important because with such a large SEM, it was difficult to identify specific hypotheses within the omnibus fit statistics. For example, we used factor scores as raw variables in the model building process. If the factor structure changes, we may have to revert to raw scores on the original measures. This has implications for the measurement of teacher behavior and is a fertile area for future investigations.

In targeting future model building, the follow-up tests we did provided guidance for respecification of the SEM based upon teaching experience, types of teaching assignment (including school level as well as in- and out-of-field status, and potential interactions between factors). Similarly, we plan to include other measures on which we have also collected data, but did not use for this exploratory model (e.g., Praxis math, reading, and writing scores). Other possible areas of study include how school-level SES can be used in conjunction with the estimated classroom diversity index we employed in Paper #2 and how these variables interact with professional development or teaching assignments. Indeed, framing teacher knowledge is a persistent challenge that we will continue to investigate, and explore with respect to subject matter knowledge via college credits in science and education, GPAs in those classes, other covariates associated with teacher backgrounds, and descriptions of their current and future teaching placements.

Lastly, further analyses will examine the specific, relative contributions of the teacher preparation program and ongoing professional development. The goal will be to determine tipping points at which either teacher preparation or teacher professional development are more effective for teachers' inquiry-based instruction. Similarly, we will examine the type of professional development as that will certainly covary with a ROI examination of teacher programs and professional development.

As states work to implement the *Next Generation Science Standards*, either adopted or adapted, they must also carefully consider if their state science teacher certification policies are supporting robust science teacher preparation. When state policymakers only require a general science base for their science teachers that renders them out-of-field in multiple areas, they risk undermining science education reform.
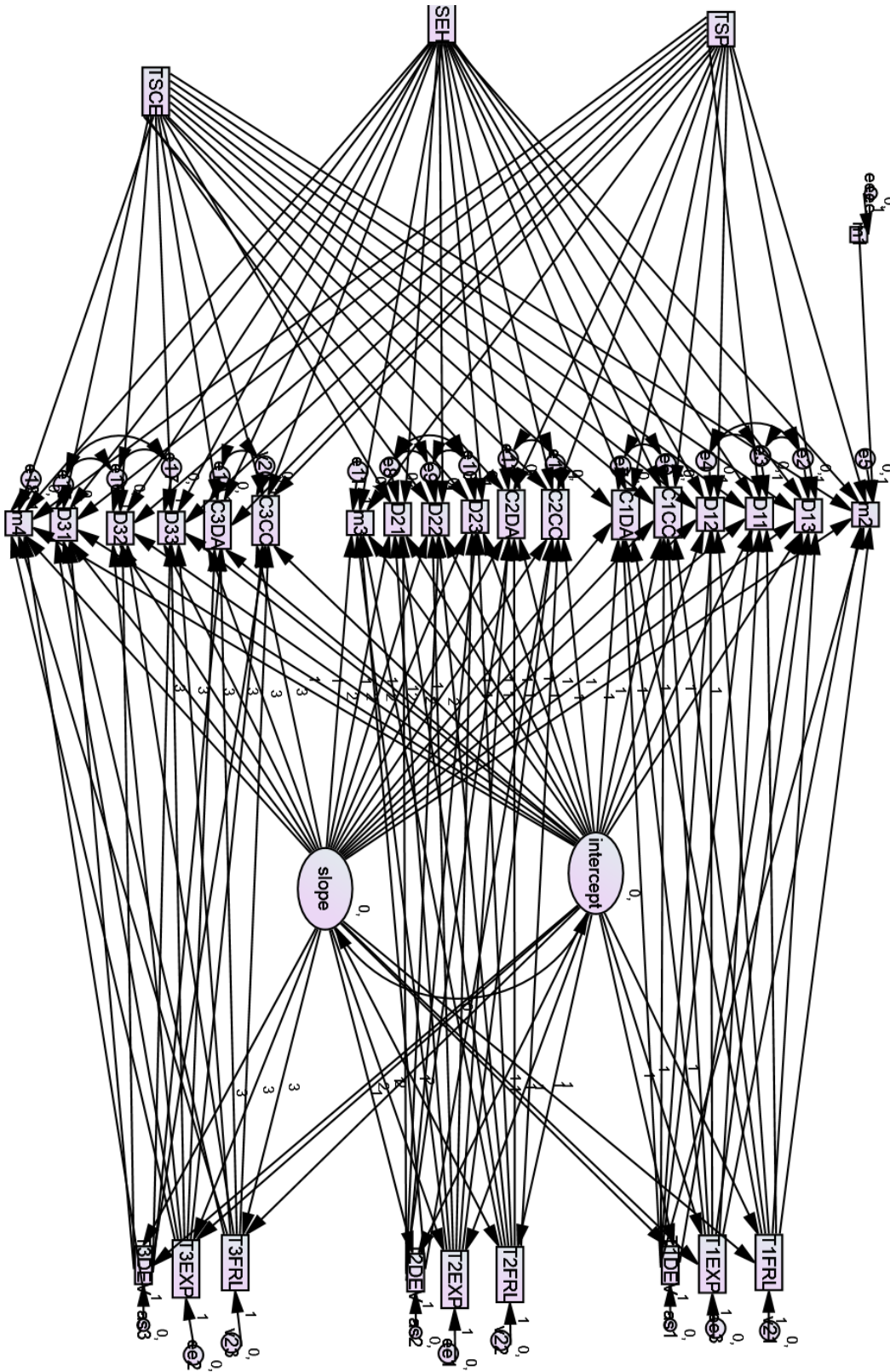
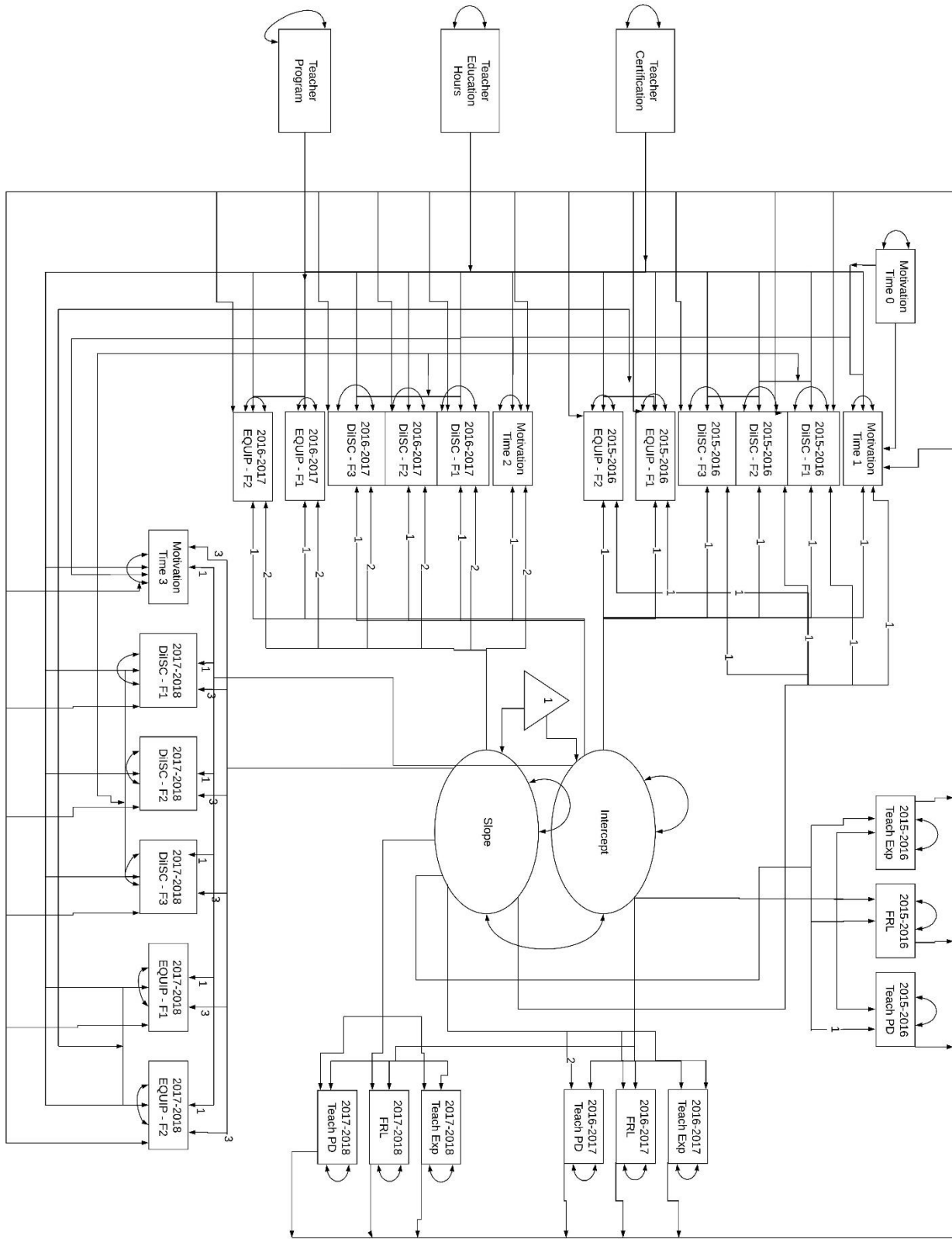*Figure 2*. Multivariate SEM from AMOS.

*Figure 3*. Simplified version of multivariate growth SEM specification.

**Acknowledgments**

**References**

Alexopoulou, E., & Driver, R. (1996). Small-group discussion in physics: Peer interaction modes in pairs and fours. *Journal of Research in Science Teaching, 33*, 1099–114.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.

Amin, T. G., Smith, C., & Wiser, M. (2014). Student conceptions and conceptual change: Three overlapping phases of research. In Lederman, N.G. Abell, S.K. (Eds.) *Handbook of Research on Science Education, Volume II*. 57-81.

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

Antink-Meyer, A., & Brown, R. A. (2017). Second-career science teachers' classroom conceptions of science and engineering practices examined through the lens of their professional histories. *International Journal of Science Education*, *39*(11), 1511–1528. https://doi.org/10.1080/09500693.2017.1338787

Anyon, J. (1980). Social class and the hidden curriculum of work. *The Journal of Education*, *162*(1), 67–92.

Baker, D., Beard, R., Bueno-Watts, N., Lewis, E., Özdemir, G., Perkins, G., Uysal, S., Wong-Kavas, S., & Yaşar-Purzer, S. (2008). *Discourse in Inquiry Science Classrooms* (DiISC): Reference Manual, Technical Report No. 001. The Communication in Science Inquiry Project (CISIP): Arizona State University.

Baker, D.R., Lewis, E.B., Purzer, S., Watts, N.B., Uysal, S., Wong, S., Beard, R., & Lang, M. (2009). The *Communication in Science Inquiry Project* (CISIP): A project to enhance scientific literacy through the creation of science classroom discourse communities. *International Journal of Environmental & Science Education, 4(3),* 259-274.

Bianchini, J. (2012). Teaching while still learning to teach: Beginning science teachers' views, experiences, and classroom practices. In B.J. Fraser, K. Tobin, & C.J. McRobbie (Eds.), *Second international handbook of science education* (pp. 389-399). The Netherlands: Springer.

Black, P. & Wiliam, D. (1998). Assessment and classroom learning in *Assessment in Education: Principles* Policy & Practice, 5 (1) (1998), pp. 7-74

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). How People Learn Brain, Mind, Experience, and School. Washington DC National Academy Press.

Brownstein, E. M., & Horvath, L. (2016). *Next Generation Science Standards and edTPA: Evidence of Science and Engineering Practices* (Vol. 20). Retrieved from https://eric.ed.gov/?id=ED571279

Budescu, D. V., & Budescu, M. (2012). How to measure diversity when you must. *Psychological Methods*, *17*(2), 215–227. https://doi.org/10.1037/a0027129

Bybee, R. (2009). The BSCS 5E Instructional Model and the 21st Century Skills. Retrieved: https://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_073327.pdf

Bybee, R. W. (2011). *The teaching of science: 21st century perspectives*. Arlington, VA: NSTA press.

Bybee, R. W. (2014). NGSS and the next generation of science teachers. *Journal of science teacher education*, *25*(2), 211-221.

Bybee, R. W., Taylor, J. A., Gardner, A., Van Scotter, P., Powell, J. C., Westbrook, A., & Landes, N. (2006). *The BSCS 5E Instructional model: Origins and effectiveness.* (pp. 88–98). Colorado Springs.

Callaghan, M., Knapp, P., & Noble, G. (1999). Genre in practice. In Cope and M. Kalantzis (Eds.). The powers of literacy: A genre approach to teaching writing (pp. 179-202). London: The Falmer Press.

Capps, D.K., Crawford, B.A. 2013. Inquiry-based professional development: what does it take to support teachers in learning about inquiry and nature of science? *International Journal of Science Education. 35(12),* 1947-1978.

Carrasquillo, A. L., & Rodriguez, V. (1996). Language minority students in the mainstream classroom. Clevedon, England: Multilingual Matters.

Carpenter, S.L., Iveland, A., Moon, S., Bianchini, J.A. (2015). Prospective science teachers' Understanding of science and engineering practices. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Chicago, IL.

Cavagnetto, A., Hand, B. M., & Norton-Meier, L. (2010). The Nature of Elementary Student Science Discourse in the Context of the Science Writing Heuristic Approach. *International Journal of Science Education*, *32*(4), 427–449. https://doi.org/10.1080/09500690802627277

Chamot, A. U., & O'Malley, J. M. (1987). The Cognitive Academic Language Learning Approach: A bridge to the mainstream. TESOL Quarterly, 21 (2): 227-249.

Childs, A., & McNicholl, J. (2007). Investigating the Relationship between Subject Content Knowledge and Pedagogical Practice through the Analysis of Classroom Discourse. *International Journal of Science Education*, *29*(13), 1629–1653. https://doi.org/10.1080/09500690601180817

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, *86*(2), 175–218. https://doi.org/10.1002/sce.10001

Crawford, B. A. (2014). From inquiry to scientific practices in the science classroom. In Lederman, N.G. Abell, S.K. (Eds.) *Handbook of Research on Science Education, Volume II* (pp. 529-556). Routledge.

Cuban, L. (1992). Curriculum stability and change. In P. W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 216-247). New York: Macmillan.

Cunningham, C. M., & Carlsen, W. S. (2014). Teaching engineering practices. *Journal of Science Teacher Education*, *25*(2), 197-210.

Czerniak, C. M., & Johnson, C. C. (2007). Interdisciplinary science teaching. *Handbook of research on science education*, 537-559.

Dalgarno, N., & Colgan, L. (2007). Supporting novice elementary mathematics teachers' induction in professional communities and providing innovative forms of pedagogical content knowledge development through information and communication technology. *Teaching and teacher education*, *23*(7), 1051-1065.

Darling-Hammond, L. (2013). *Getting teacher evaluation right: What really matters for effectiveness and improvement*. Teachers College Press.

Darling-Hammond, L. (2016). Research on Teaching and Teacher Education and Its Influences on Policy and Practice. *Educational Researcher*, *45*(2), 83–91. https://doi.org/10.3102/0013189X16639597

Dewey, J. (1938). Experience and education.

Drew, S. V., & Thomas, J. (2018). Secondary science teachers' implementation of CCSS and NGSS literacy practices: a survey study. *Reading and Writing*, *31*(2), 267–291. https://doi.org/10.1007/s11145-017-9784-7

Ericsson, K. A., & Simon, H. A. (1998). How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. Mind, Culture, and Activity, 5(3), 178-186.

Fradd, S. H., & Lee, O. (1999). Research news and comment: Teachers' roles in promoting science inquiry with students from diverse language backgrounds.

Ford, M. J. (2015). Educational implications of choosing "practice" to describe science in the Next Generation Science Standards. *Science Education*, *99*(6), 1041-1048.

French, D. A., & Burrows, A. C. (2018). Evidence of Science and Engineering Practices in Preservice Secondary Science Teachers' Instructional Planning. *Journal of Science Education and Technology*. https://doi.org/10.1007/s10956-018-9742-4

Halliday, M., & Martin, J. (1993). Writing science: Literacy and discursive power. Pittsburgh, PA: University of Pittsburgh Press.

Hashweh, M. Z. (1987). Effects of subject-matter knowledge in the teaching of biology and physics. Teaching and Teacher Education, 3(2), 109-120.

Herrell, A. L., & Jordan, M. (2004). Fifty strategies for teaching English language learners (2nd ed.). Upper Saddle River, NJ: Merrill.

John-Steiner, V., & Mahn, H. (2012). Sociocultural Approaches to Learning and Development-A Vygotskian Framework. *Educational Psychologist*, *31*. https://doi.org/10.1080/00461520.1996.9653266

Jonassen, D. H. (1991). Objectivism versus constructivism: Do we need a new philosophical paradigm? *Educational technology research and development*, *39*(3), 5-14.

Kane, M. (1990). An argument-based approach to validation. ACT Research Report Series 90-13. Iowa City, Iowa: The American College Testing Program.

Kelly, G. J. (2014). Discourse practices in science learning and teaching. In Lederman, N.G. Abell, S.K. (Eds.) *Handbook of Research on Science Education, Volume II*. (321-336). New York, NY: Routledge.

Kelly, G. J., & Crawford, T. (1997). An ethnographic investigation of the discourse processes of school science. Science Education, 81, 533–559.

Kelly, G. J., & Green, J. (1998). The social nature of knowing: Toward a sociocultural perspective on conceptual change and knowledge construction. In B. Guzzetti & C. Hynd (Eds.), Perspectives on conceptual change: Multiple ways to understand knowing and learning in a complex world (pp. 145–181). Mahwah, NJ: Erlbaum.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational psychologist*, *41*(2), 75-86.

Kittleson, J. M., & Southerland, S. A. (2004). The role of discourse in group knowledge construction: A case study of engineering students. Journal of Research in Science Teaching, 41, 267 – 293.

Klein, P. (1999). Learning science through writing: The role of rhetorical structures. The Alberta Journal of Educational Research, XLV (2), 132-153.

Kloser, M. (2014). Identifying a core set of science teaching practices: A delphi expert panel approach: CORE SCIENCE TEACHING PRACTICES. *Journal of Research in Science Teaching*, *51*(9), 1185–1217. https://doi.org/10.1002/tea.21171

Kvale, S., & Brinkmann, S. (2009). Interviews: Learning the craft of qualitative research interviewing.

Lave, J., & Wenger, E. (1992). *Situated learning: Legitimate peripheral participation*. Cambridge university press.

Lee, O. (2017). Common core state standards for ELA/literacy and next generation science standards: Convergences and discrepancies using argument as an example. *Educational Researcher*, *46*(2), 90-102.

Lee, O. & Fradd, S.H. (1998). Science for all, including student from non-English language backgrounds. New York: Cambridge University Press.

Lemke, J. (1990). Talking science: Language, learning, and values. Norwood, NJ: Ablex.

Lemke, J. L. (2001). Articulating communities: Sociocultural perspectives on science education. *Journal of Research in Science Teaching*, *38*(3), 296–316. https://doi.org/10.1002/1098-2736(200103)38:3<296::AID-TEA1007>3.0.CO;2-R

Lewis, E.B., Baker, D., Bueno Watts, N., & van der Hoeven Kraft, K. (2016). Chapter 1: Science Teachers' Professional Growth and the Communication in Science Inquiry Project. In *Professional Development: Recent Advances and Future Directions.* Hauppauge, NY: Nova Science Publishers, Inc., pp. 1-57.

Lewis, E.B., Baker, D.R., & Helding, B.A. (2015). Science teaching reform through professional development: Teachers' use of a scientific classroom discourse community model. *Science Education, 99(5),* 896-931.

Lewis, E.B., Rivero, A., Musson, A., Lucas, L., Tankersley, A., & Helding, B.A. (2019). Chapter 4: Educating Effective Science Teachers: Preparing and Following Teachers into the Field. In J. Carinci, S. Meyer, and C. Jackson (Eds.), *Linking Teacher Preparation Program Design and Implementation to Outcomes for Teachers and Students.* Charlotte, NC: Information Age Publishing.

Lewis, E.B., Rivero, A., Lucas, L., Tankersley, A., & Helding, B. (2018). Beginning Science Teachers' Subject Matter Knowledge, Misconceptions, and Emerging Inquiry-based Teaching Practices. Paperset presented at the annual meeting of the National Association for Research in Science Teaching (NARST): March 10-13, 2018: Atlanta, Georgia.

Lucas, L., & Lewis, E. B. (2017). Modeling Inquiry-oriented Instruction of Beginning Secondary Science Teachers.

Luft, J. A., Firestone, J. B., Wong, S. S., Ortega, I., Adams, K., & Bang, E. (2011). Beginning secondary science teacher induction: A two-year mixed methods study. *Journal of Research in Science Teaching*, *48*(10), 1199–1224. https://doi.org/10.1002/tea.20444

Kane, M. (1990). An argument-based approach to validation. ACT Research Report Series 90-13. Iowa City, Iowa: The American College Testing Program.

Marshall, J.C., Horton, B., Smart, J, & Llewellyn, D. (2008). *EQUIP: Electronic Quality of Inquiry Protocol:* Retrieved from www.clemson.edu/iim.

Marshall, J.C., Smart, J. & Horton, R.M (2010). The design and validation of EQUIP: An instrument to assess inquiry-based instruction. International Journal of Science and Mathematics Education, *8*(2), 299-321.

Messick, S. (1987). Validity. ETS Research Report Series, 1987(2), i-208.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American psychologist, 50(9), 741.

Miles, M. B., Huberman, A. M., Huberman, M. A., & Huberman, M. (1994). Qualitative data analysis: An expanded sourcebook.

Moje, E., Collazo,T, Carilo, R., & Marx, R. (2001). Maestro, what is "quality?": Language, literacy and discourse in project-based science. Journal of Research in Science Teaching, 38, 469-498.

Morales, C. J. (2016). *Adapting to National Standards: The experience of one middle school science teacher's implementation of the Next Generation Science Standards (NGSS)*. University of Michigan. Retrieved from https://deepblue.lib.umich.edu/handle/2027.42/133382

National Research Council (1996). National science education standards. Washington, DC:

National Research Council (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington, D.C.: National Academies Press.

National Research Council (2005). How students learn: History, mathematics and science in the classroom, a targeted report for teachers, M Donovan and J. Branford (Eds.). Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academy Press.

National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K-8*. National Academies Press.

National Research Council (2010). *Preparing teachers: Building evidence for sound policy*. Washington, D.C.: National Academies Press.

National Research Council (2012). A framework for K-12 science education: Practices, cross- cutting concepts, and core ideas. Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Science and Education. Washington, DC: The National Academies Press.

Newton, P. Driver, R. & Osborne, J. (1999). The place of argumentation in the pedagogy of school science. International Journal of Science Education, 21(5), 553-576.

NGSS Lead States (2013). Next Generation Science Standards: For States, By States. Achieve, Inc. on behalf of the twenty-six states and partners that collaborated on the NGSS.

Nixon, R. S., Luft, J. A., & Ross, R. J. (2017). Prevalence and predictors of out-of-field teaching in the first five years. *Journal of Research in Science Teaching, 54*(9), 1197-1218.

Oakes, J., & Guiton, G. (1995). Matchmaking: The dynamics of high school tracking decisions. *American Educational Research Journal*, *32*(1), 3–33.

Porcaro, D. (2011). Applying constructivism in instructivist learning cultures. *Multicultural Education & Technology Journal*, *5*(1), 39-54.

Rivard, L. P. (1994). A review of writing to learn in science: Implications for practice and research. Journal of Research in Science Teaching, 31, 969-983. doi:10.1002/tea.3660310910

Rivard, L. P., & Straw, S. B. (2000). The Effect of Talk and Writing on Learning Science An Exploratory Study. Science Education, 84, 566-593.

Roth, W.-M., & Bowen, G. M. (1995). Knowing and Interacting: A Study of Culture, Practices, and Resources in a Grade 8 Open-Inquiry Science Classroom Guided by a Cognitive Apprenticeship Metaphor. *Cognition and Instruction*, *13*(1), 73–128.

Sadler, P., & Sonnert, G. (2016). Understanding misconceptions: Teaching and learning in middle school physical science. *American Educator, 40*(1):26–32.

Sandoval, W. A., Kawasaki, J., Cournoyer, N., & Rodriguez, L. (2016). *Secondary Teachers' Emergent Understanding of Teaching Science Practices*.

Scantlebury, K. (2014). Gender matters: Building on the past, recognizing the present, and looking toward the future. In *Handbook of Research on Science Teaching* (Vol. 2). Taylor & Francis.

Schneider, R., Plasman, K. (2011). Science teacher learning progressions: A review of science teachers' pedagogical content knowledge development. *Review of Educational Research. 81*(4), 530-565.

Sizer, T. R. (1984). *Horace's Compromise: The Dilemma of the American High School: with a New Preface*. Boston, MA: Houghton Mifflin Harcourt.

Summers, R. L., Jackson, L., Woodward, L. H., Jones, J., & Dryer, S. (2011). Methodology for measuring diversity in human populations. *Journal of Cultural Diversity*, *18*(2).

Tam, M. Y. S., & Bassett, G. W. (2004). Does diversity matter? Measuring the impact of high school diversity on freshman GPA. *Policy Studies Journal*, *32*(1), 129–143.

Taylor, P.C. (2014). Contemporary qualitative research: Toward and integral research perspective. In Lederman, N.G. Abell, S.K. (Eds.) *Handbook of Research on Science Education, Volume II.* (38-56). New York, NY: Routledge.

Treagust, D. F., & Tsui, C. Y. (2014). General instructional methods and strategies. In N.G. Lederman & S.K. Abell (Eds.), *Handbook of Research on Science Education Volume II,* (pp. 303-320). New York, NY: Routledge.

Tschannen-Moran, M. & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*, 783-805.

Wenger, E. (1998). Communities of practice: Learning as a social system. *Systems thinker*, *9*(5), 2-3.

Windschitl, M. (2004). Caught in the cycle of reproducing folk theories of "Inquiry": How pre-service teachers continue the discourse and practices of an atheoretical scientific method. Journal of Research in Science Teaching, 41(5), 481 – 512.

Woodruff, E., & Meyer, K. (1997). Explanations from intra-and inter-group discourse: Students building knowledge in the science classroom. *Research in Science Education*, *27*(1), 25–39.

Yore, L.D., Hand, B., & Prain, V. (January, 1999). Writing-to-learn science: Breakthroughs, barriers, and promises. Paper presented at the International Conference of the Association for Educating Teachers in Science, Austin, TX.