Spring 7-26-2017

Using the Task Analysis Process with Teachers to Uncover Language Demands within an Eight-Week NGSS Summer Course

Leah Plack
Portland State University

Let us know how access to this document benefits you.

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds

Part of the Elementary Education Commons, and the Science and Mathematics Education Commons

Recommended Citation

Using the Task Analysis Process With Teachers to Uncover Language

Demands Within an Eight-Week NGSS Summer Course

by

Leah Plack

A thesis submitted in partial fulfillment of the
requirements for the degree of

Master of Science in Teaching
in
General Science

Thesis Committee:
William Becker, Chair
Stephanie Wagner
Susan Holveck

Portland State University
2017

Abstract

The state of Oregon has adopted the Next Generation Science Standards as well as the English Language Proficiency standards, both of which affect elementary school teachers. These standards adoptions challenge teachers and professional developers to rethink how they approach science conceptual understanding and language acquisition. The Math Science Partnership K-6 Instructional Specialist Grant made the decision to incorporate a Task Analysis process, which asks the participant to analyze the demands of a content-based task in the domains of content knowledge, analytical skills and language, into six eight-week summer courses focused on the Next Generation Science Standards. A pre and post-assessment was created to determine if any growth in teacher understanding of the demands of a science task could be observed as a result of engaging in the Task Analysis process. A four point rubric was created to score participant responses. Two research questions were developed: 1. How well does the ELPA21 Task Analysis tool help participants understand the language demands of a science task when used as part of an NGSS summer professional development course? 2. How effective is a work sample and scoring rubric protocol for measuring the impact of the Task Analysis process? Participants showed statistically significant growth in their analysis of a science task from pre to post-assessment responses, with particularly strong growth demonstrated in the domains of content and language. Further coding of responses revealed that participants frequently discussed vocabulary as both a language and content knowledge demand of a science task.

Table of Contents

List of Tables

List of Figures

Introduction

Across the country, many school districts are observing the same trends: the numbers of English learners (ELs) are increasing, while the numbers of English as a Second Language instructors are decreasing (Santos et. al, 2012). In the state of Oregon, the number and percentage of ELs continue to rise, while a significant gap in school achievement, graduation rates and college completion remains between ELs and their English proficient peers (ODE, 2013). It is becoming less and less feasible to hand off the teaching of English to ESL teachers, and increasingly, the charge of instructing our students in the English language is being placed on the general classroom teacher (CCSSO, 2012). In Oregon, ELs lag far behind their peers in content area assessments: in 2015-16, 56.5% of all fifth grade students met or exceeded the proficiency benchmark in English language arts, while only 12.8% of students designated as Limited English Proficient (LEP) met or exceeded the benchmark (ODE, 2016). 40.4% of all fifth grade students in Oregon were proficient in math, while only 6.5% of LEP students were designated as proficient. In science, 58.5% of all fifth grade students were proficient, while only 18% of LEP students were designated as proficient. It is clear that students who speak English as a second language in Oregon are not performing on standards-based assessments at the level of their English proficient peers.

Simultaneously, Oregon's adoption of the the Next Generation Science Standards (NGSS) is placing more emphasis on understanding nonfiction text, arguing from evidence and constructing explanations of phenomena, among other practices,

which require all students to use language in science (Cheuk, 2013; Lee, Quinn, &

Valdés, 2013). These practices pose great challenges to English learners if they are not

accompanied by language supports (Lee, Quinn, & Valdés, 2013). If proper supports are

in place, English learners will be afforded opportunities to use language in authentic

ways within the science classroom. In order to capitalize on these opportunities for

authentic language that accompany the NGSS, it is essential that teachers understand

the language demands these standards place on English learners, so that they can

provide the necessary scaffolds to help English learners perform the daunting task of

learning English while mastering content (CCSSO, 2012).

The implementation of the Next Generation Science Standards brings to light

many ongoing challenges and barriers identified by elementary teachers for teaching

science. Elementary teachers commonly lack confidence in their understanding of

science content and in teaching science (Banilower et al., 2013). Time and material

resources necessary for science are often limited or nonexistent in elementary schools.

Furthermore, with the high stakes testing of the No Child Left Behind initiative, many

elementary teachers have been pressured into increasing instructional minutes for

English language arts and math, at the expense of science and other subjects

(McMurrer, 2008). With the implementation of the NGSS, teachers are asked to master

new science content as well as what may be novel pedagogy, often with limited support.

In light of these language demands presented by the NGSS, the inherent

challenges in teaching the NGSS, the excitement of new opportunities for authentic

language for ELLs, and the shifts in English language development instruction called

for by the ELP standards, the question looms of how to best support teachers to be prepared to usher in these changes. There is very little research on how best to roll out the NGSS and ELP standards, as the standards are still so new, but there is a body of research that focuses on the implementation of previously adopted standards that can help guide future efforts (Penuel, Harris & DeBarger, 2015).

In this particular context, I studied the efforts of the Math Science Partnership (MSP) K-6 NGSS Instructional Specialist Program (known as the "MSP grant"), in partnership with Portland State University, to develop science teacher leaders in the Portland Metro Area. The goal of this program is to provide in-depth training on the NGSS to a cohort of 74 elementary and middle school teachers, so they may be empowered to improve their own classroom practice, and become science leaders at their schools and within their districts. The main foci of the Teacher Leadership Program are to increase teacher science content knowledge, pedagogical content knowledge, curricular content knowledge, and leadership capacity. Within the umbrella of pedagogical content knowledge, the MSP grant leadership specifically expressed concern that more needed to be done to prepare teacher participants to teach NGSS-aligned science to English learners.

Oregon has adopted both the Next Generation Science Standards and the English Language Proficiency standards. The MSP Instructional Leadership Team felt it was important to ensure meaningful coursework for teacher participants by tying any language discussions to the shifts in how language should be taught, as guided by the ELP standards. With this goal in mind, the Instructional Leadership Team discovered

3

and began using the Task Analysis resource found within the ELPA21 professional development resources. This tool focuses on analyzing a task in any content area to determine what students need to know and be able to in the domains of content knowledge, analytical skills and language. The tool was designed to be used collaboratively to facilitate meaningful conversations about what everyday content-based tasks are asking of English learners (ELs) in the classroom. These discussions can serve as the first step towards discussions of how to appropriately scaffold tasks for ELs so that they may fully participate in content-based classroom experiences.
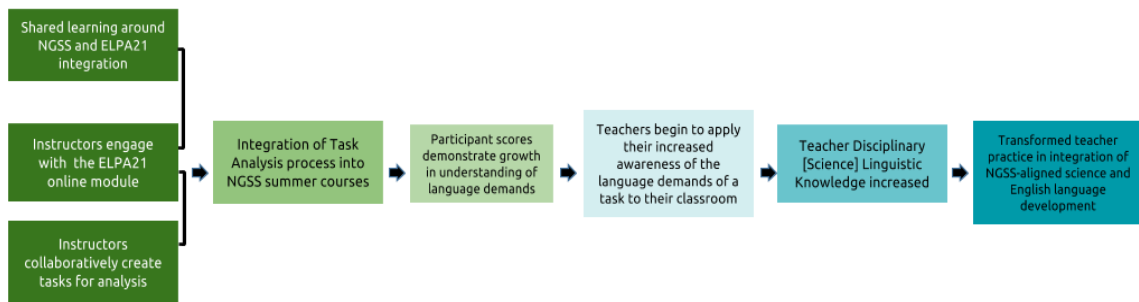
Figure 1: Theory of Change

As part of the MSP Instructional Specialist grant's second year program design, three summer courses were offered under the organizing themes of three Crosscutting Concepts: Energy and Matter, Change Over Time, and Interactions and Systems. The Task Analysis process was selected as the treatment that would be integrated into the courses, the effects of which would be studied. The Theory of Change (see Figure 1) details the desired outcomes of this treatment. It should be noted that this study only measured results from the first three steps of the Theory of Change; classroom outcomes are the eventual goal but determining change in teaching practice was

outside the scope of this study. While observing the preparation for and outcomes of these NGSS courses, as well as developing the Task Analysis rubric, we asked the following questions: 1. How well does the ELPA21 Task Analysis tool help participants understand the language demands of a science task when used as part of an NGSS summer professional development course? 2. How effective is a work sample and scoring rubric protocol for measuring the impact of the Task Analysis process?

In order to answer these questions, an assessment was created using the questions from the Task Analysis tool. This assessment was designed to capture the dependent variable of change in understanding of the Task Analysis process from the onset of the course to the final session. Along with the pre and post-assessments, a rubric was designed to rate responses on a four point scale in order to quantify any change in understanding that we observed. Participant scores were then compared from pre to post-assessments, and the comparisons were broken down by several different factors including which course the participant was enrolled in, whether they were part of the cohort or at large population, and how long they had been teaching. Participants and course instructors were interviewed about their experiences.

Literature Review

With the adoption of the Next Generation Science Standards and the English Language Proficiency Standards, the K-6 Math Science Partnership grant staff tasked themselves with the challenge of incorporating a language development component into their NGSS summer course programming. To better understand the "whys" and "hows" of developing this new aspect of the courses, it was necessary to current recommendations around the shifts in science and language teaching practice that would be necessary to address the new standards, current understandings of recommended teacher pedagogical content knowledge in science and language, and studies of teacher professional development programs that focused on inquiry science and language.

Shifts in Science and Language Teaching Practice with NGSS and ELP Standards

Oregon has adopted the Next Generation Science Standards (NGSS), which are national K-12 standards meant to guide districts and schools in providing comprehensive science instruction so that students graduate scientifically literate and ready for college and career (The Need for Standards, n.d.). The NGSS are "Three Dimensional" standards, meaning that there are three components that interweave to create performance expectations for students (NRC, 2011). The three components are Disciplinary Core Ideas (DCIs), Science and Engineering Practices (SEPs) and Crosscutting Concepts (CCs). Teaching in a "three dimensional" way means that teachers must be very familiar with all three dimensions in order to interweave them into their lesson plans and assessments. This is a very new approach for most science teachers, because previous

Oregon standards separated practices (referred to as "science process skills") from content and did not explicitly emphasize Crosscutting Concepts (ODE, 2009).

Okhee Lee, Helen Quinn, and Guadalupe Valdés (2013) discuss the implications of the NGSS for language development, and specifically for English learners, in their paper Science and Language for English Language Learners. In this paper, Lee, Quinn and Valdés emphasize the features of the NGSS which are particularly language intensive, as well as summarizing the literature base around science and language acquisition, with recommendations for future research and policy.

For each dimension of the NGSS, the ability to access and use the required language is critical for students to make meaning science, and to communicate to others what they have learned. The DCIs, or science and engineering content that students should be able to demonstrate understanding of, are full of content specific language and terms. The Science and Engineering Practices (SEPs) are a set of behaviors and skills engaged in by scientists, as well as ways of constructing meaning in the science classroom (Lee, Quinn & Valdés, 2013). All SEPs involve language to some degree with four practices highlighted for being the most language intensive: Arguing From Evidence, Obtaining, Communicating and Evaluating Information, Developing and Using Models, and Engaging in Arguing From Evidence. The Crosscutting Concepts are concepts that appear in every discipline of science, such as Stability and Change, and Patterns. When teachers make the CCs explicit to students, students connect what they are learning to what they have previously learned, and by doing so, strengthen their conceptual understanding (NRC, 2011). Lee, Quinn and Valdes (2013) argue that with

the proper supports in place, as student scientific conceptual understanding grows, so too will their use of language.

Lee, Quinn and Valdés (2013) advocate for a shift towards *"language of the science classroom"* (p. 228)*.* This includes moving away from content-based language instruction and the sheltered model, in which language forms and functions are emphasized within a science context, and towards a model of language acquisition that happens while students, with appropriate supports in place, are engaging in doing, discussing and making meaning of science with their peers. Lee, Quinn and Valdés (2013) also recommend a shift in thinking away from "academic language," which too often is synonymous with content specific vocabulary and language forms, and may overly emphasize student deficits, and towards "science registers," which describe the different types of language teachers and students may engage in within a science classroom, ranging from colloquial to discipline-specific.

Lee, Quinn and Valdés (2013) conclude that the NGSS provide many rich opportunities for all students to use language, and that it is critical that ELs be afforded these same opportunities. They recommend that teachers focus less on expecting perfect language from ELs, and more on encouraging equitable participation and sharing of scientific ideas. The authors pose several future research questions that will need to be addressed, including: "What do science teachers and language specialists need to know about language demands and opportunities to support ELLs' engagement in these practices (p. 231)?"

In addition to the shifts occurring in science education, thinking around how to teach English to English language learners has also shifted. Previously, English Language Development (ELD) instruction, driven by the Oregon English Language Proficiency standards, focused greatly on vocabulary development and the forms of language (CTB/McGraw-Hill, 2007). English instruction was to be focused primarily on English conventions, and students learning English would not be considered proficient until they could speak with "native speaker abilities" (CTB/McGraw-Hill, 2007, p.2). As a consequence, authentic opportunities to learn language and content simultaneously were not the focus of ELD instructional time, and were in fact often discouraged.

With the shifts in what the NGSS are asking students to do with language and content, new thinking around how to teach ELD was also needed. The Understanding Language Initiative at Stanford University was formed to call attention to the needs of English language learners within the CCSS and NGSS. In order to organize their future work, three major shifts in teaching English language development were identified:

1. Language acquisition, rather than an individual cognitive process, is a social process through which language is acquired in social contexts.

2. The acquisition process, rather than involving primarily the sequenced building of forms and structures and vocabulary aimed at accuracy, fluency and complexity, is a non-linear and complex developmental process aimed at comprehension and communication.

3. Participation in activity simultaneously develops conceptual understanding and language use. (CCSSO, 2012, p.43)

These shifts represent a movement towards authentic language that is used in social, academic contexts. While language forms and vocabulary are certainly still considered important, they are no longer the driving force behind student language learning and assessment. Now at the forefront is students' ability to express their thinking to others, to collaboratively build conceptual understanding through use of language, and to receive information through reading and listening. In addition, much greater emphasis is placed on connecting all classroom learning to ELLs' home language and culture, and utilizing their "funds of knowledge" within the classroom (Moll et al., 1992).

Brown and Ryoo (2008) expand on this idea of moving away from the emphasis on content specific vocabulary. In their study, they worked with a diverse elementary school in California to provide science instruction to 49 5th grade students. Half of the students were randomly assigned to the treatment group, and half to the control group. Web-based lessons were used to remove the variable of teacher implementation of the treatment program.

Two versions of a science educational program were created and administered: an "aggregate" program (control group), which focused on teaching science concepts and content-specific vocabulary simultaneously, and a "disaggregate" or "content-first" program, which used everyday language to teach science concepts and reinforce content-specific language (treatment group) (Brown & Ryoo, 2008). The theoretical assumption for the disaggregate program was that students would be better primed to learn the scientific vocabulary later if they first understood the science concepts. Both programs were taught in three phases: the content construction phase, where

10

students were introduced to scientific concepts, the explicit language phase, where

students were taught the language needed to explain the scientific phenomenon, and

the scaffolding opportunities for language phase, where students had an opportunity to

demonstrate their understanding by using experiments to answer questions about the

phenomenon. In the content construction phase, the control group was taught the

academic vocabulary necessary to describe a concept using everyday language,

subsequently, only the vocabulary word was used (ex: "glucose" was defined as "a type

of sugar that plants use;" it was only referred to as glucose moving forward in the unit).

The treatment group was not yet presented with the academic terminology, but

instead, was taught about the scientific concepts using only everyday language. In the

explicit language phase, students in the control group were only presented information

in academic language, while the treatment group was presented with both academic

and everyday language and asked to use both to discuss the concepts. The scaffolding

opportunities for language phase was presented identically to both the treatment and

the control groups of students.

The study used a pre-post control group design with three dependent measures:

overall scores, aggregate scores (which scored students' ability to understand and

explain concepts using academic language) and disaggregate scores (which scored

students' ability to understand and explain concepts using everyday language) (Brown &

Ryoo, 2008). Assessments used multiple choice questions as well as open ended

questions which asked students to explain the phenomenon of photosynthesis. The

results were analyzed using an independent *t* test and Cohen's d to determine effect

size. Results showed near identical pre-test results, and much higher gains in post-test results from the treatment group, with a mean of 17.8 points improvement as opposed to 8.5 for the control group. The improvements of the treatment group were statistically significant, with a large effect size shown. Furthermore, students in the treatment group showed gains in both science understanding and use of scientific language. The gains for the treatment group in use of scientific language were especially significant. This study also shed light on how we assess student understanding in science - are assessments informing educators about conceptual understanding in science, or how well students understand and use scientific language? Brown and Ryoo's study highlights the need to consider language demands as part of science instruction and assessment, which is precisely what the English Language Proficiency standards are designed to do.

The English Language Proficiency standards that Oregon has adopted were developed through a collaboration of the Stanford Understanding Language initiative and WestEd, an educational consulting firm. The ELP standards connect to content area standards, and are meant to support deep connections between language and content learning for both students and educators (CCSSO, 2014). The Understanding Language Initiative, key in writing the ELP standards, created a Task Force whose role was to examine currently implemented English Language Development standards and the processes that led to their development. In their report, the Task Force had this to say (Understanding Language Initiative, 2012):

At present, second language development is seen largely as the responsibility of the ESL/ELD teacher, while content development as that of the subject area

teacher. Given the new [content] standards' explicitness in how language must

be used to enact disciplinary knowledge and skills, such a strict division of labor

is no longer viable. Content area teachers must understand and leverage the

language and literacy practices found in science, mathematics, history/social

studies, and the language arts to enhance students' engagement with rich

content and fuel their academic performance. ESL/ELD teachers must cultivate a

deeper knowledge of the disciplinary language that ELL students need, and help

their students to grow in using it. Far greater collaboration and sharing of

expertise are needed among ESL/ELD teachers and content area teachers at the

secondary level. At the elementary level, far greater alignment and integration

are needed across ESL/ELD and subject matter learning objectives, curriculum,

and lesson plans that teachers in self-contained classrooms prepare and deliver.

(p. 2)

Walqui and Heritage (2012) from Stanford University's Understanding Language

Initiative also propose that language acquisition cannot be taught in isolation from

content. When considering students who are English language learners, it is important

to remember that they represent a diverse group with different needs. Teachers, in

turn, need to employ diverse strategies and approaches to meet their needs, as well as

honoring the "funds of knowledge" that students bring from their prior experience and

home culture (Moll et. al, 1992). Walqui and Heritage (2012) argue that every teacher is

responsible for teaching English as well as content area literacy to all students.

Instruction should be based on five principles that have been formed through the literature base as well as the direct experience of the researchers.

"Principle 1: Learning is always based on prior knowledge and experience. ELLs must have equal access to knowledge that is valued in school (Walqui & Heritage, 2012, p. 1)." Every student brings their prior knowledge to school, regardless of home language, socioeconomic status or home culture. All academic learning happens within a context that is assumed to be shared between the student, teacher, and the writer of the text or designer of the experience. If students lack that context, they are automatically at a disadvantage, both in understanding the academic language and understanding the content. Commonly, this issue is combatted by simplifying the language and the content that ELLs engage with, however, this approach stifles growth in language and conceptual understanding. Teachers of ELLs must therefore weigh text and tasks for language demands as well as content complexity to find the right fit for their students. It may also be necessary to spend time building students' background knowledge so that they are able to "fill in the gaps" in order to understand the new content that they are presented with.

"Principle 2: Language and cognition develop together and progressively. As ideas and relationships become more complex, so does language (Walqui & Heritage, 2012, p. 2)." Students require language to code information and then later retrieve it as part of their conceptual understanding (Walqui & Heritage, 2012). Conceptual understanding and language are equally important, and should be learned in an integrated way through meaningful experiences and with appropriate scaffolding (as

cited in van Lier, 2004; Heritage, Silva & Pierce, 2007). This approach requires that

teachers consider their content and language goals simultaneously, as well as how these

goals support one another (Walqui & Heritage, 2012).

"Principle 3: The goal of learning is to develop the stance of generativity and

autonomy. This is accomplished through apprenticeship in which the learner is invited

to become a member of a community of practice (Walqui & Heritage, 2012, p. 3)." In

order to foster independence of ELL students, teachers need to create communities of

practice where students can participate in the established social practices of the

classroom (as cited in van Lier, 2004). Within these communities of practice, ELLs learn

language and content as apprentices. Norms are created as a group and followed, and

expectations are clear. The teacher's role in ensuring participation of all students is

crucial, as is student acquisition of the required "linguistic resources" needed to

participate (as cited in Block, 2003; Firth & Wagner, 1997; Lantolf & Thorne, 2006).

Teachers also need to scaffold activities, and to provide consistency in the structure and

rules of activities, so that students have the foundation needed in order to innovate

(Walqui & Heritage, 2012). In this way, students view themselves as members of a

community of practice with agency and autonomy.

"Principle 4: The goal of language use is to make it contextually appropriate;

students need to be competent navigators within a range of different registers (Walqui

& Heritage, 2012, p. 4)." Skilled language users understand both the concept of

registers, which are ways of communicating associated with a specific context, and

genres, which are ways of communicating that have familiar patterns of organization, 15

and how to use registers and genres appropriately. In order for teachers to effectively teach students how to navigate the use of register and genre, they themselves must first understand these concepts, as well as how to help students understand. When assessing productive language, emphasis should first be placed on student understanding of the purpose of communication, followed by their recognition of genre, then how ideas are organized, and finally on correctness of language use.

"Principle 5: Assessment is integrated into the process of teaching and learning. Assessment-elicited information is used by both teachers and students to consistently keep learning moving forward (Walqui & Heritage, 2012, p. 5)." Assessment needs to answer the questions "Where am I going? Where am I now? Where to next?" (p. 5, as cited in Hattie & Timperley, 2007; Sadler, 1989). Teachers identify short term language learning goals for their students, gather evidence about their students' current learning, and interpret that data so that the teacher and students can make decisions together about what comes next. In this way, assessment is interwoven into learning activities, as opposed to something that exists separately from the learning (Walqui & Heritage, 2012).

Pedagogical Content Knowledge in Science and Language

The aforementioned shifts in language and science instruction ushered in by the new standards point to a great need for teacher professional development (Lee, Quinn & Valdés, 2013; Lee, Miller, & Januszyk, 2014; Walqui & Heritage, 2012). With the adoption of the CCSS in ELA and Math and the NGSS, as well as the ELP standards, teachers must be prepared to change both what they are teaching and how they are

teaching it. When considering teacher practice, it is helpful to examine the construct of Pedagogical Content Knowledge (PCK). Pedagogical Content Knowledge is defined as "the knowledge that teachers use in transforming subject matter knowledge into forms that are comprehensible to students" (Park et al., 2011, p. 246). When Shulman introduced the term of Pedagogical Content Knowledge (1986), he was attempting to define the knowledge necessary for teaching that goes beyond content knowledge.

Park, Jang, Chen and Jung (2011) provide evidence for the connection between Pedagogical Content Knowledge and reform-based instruction in science. They define reform-based instruction as being based in constructivist teaching, where knowledge is not transferred directly but instead is constructed by the student based on their own experiences in the classroom, and standards-based. This description of reform-based teaching mirrors the instructional shifts required of the NGSS. PCK is defined as teacher understanding of student misconceptions and learning difficulties, and the instructional strategies used to address them. The researchers hypothesized that teachers who scored higher in their PCK would also score higher in reform-based instruction. A secondary focus of the study was to examine the use of two instruments to measure these constructs: a PCK rubric developed by Park, Chen, & Jang (2008), and the Reformed Teaching Observation Protocol (RTOP), designed to measure reform-based instruction, developed by Sawada et al. (2002).

Park et al. (2011) videotaped 7 high school biology teachers through the course of the study. A team of science education graduate students used the RTOP rubric to assess the lessons for reform-based characteristics, and then used the PCK rubric to

assess the same videotaped lessons for evidence of teacher pedagogical content knowledge. The two sets of results were compared using Pearson correlation coefficients. Through comparing the results, the researchers found that there was indeed a strong correlation between PCK and reform-based instruction. The highest correlation they found was between procedural knowledge, or scientific reasoning, and PCK. This finding implies that teachers who understand student scientific reasoning are also highly capable of identifying student misconceptions, and implementing teaching strategies that address them. A second strong correlation was between the Lesson Design and Implementation construct of the RTOP and PCK, which the researchers expected because the Lesson Design section of the RTOP includes many elements of PCK, such as the consideration of students' prior knowledge and preconceptions.

Park et al. (2011) concluded that their research showed a strong link between PCK and reform-based instruction. Therefore, they argue that when discussing reform in science education, we are also referring to PCK. It is the hope of the authors that by connecting PCK to science education reform, they may be paving the way for commonality in defining what future science classrooms could and should look like. In addition, they have determined that PCK is a "reliable predictor of what a teacher knows and what the teacher is actually doing in the classroom" (Park et. al, 2011, p. 253-4). Therefore, increasing teacher PCK should be a goal of any science professional development.

As new standards are introduced, the Pedagogical Content Knowledge needed to teach effectively is undoubtedly shifting along with them. Therefore, when thinking

about In the context of the ELP standards and NGSS, it is helpful to articulate what PCK might mean within Oregon's present and future context. To define this, we look to the position papers of the ELP standards and the NGSS. To successfully support ELLs in language acquisition, teachers will need to be equipped to address language progressions, language demands, language scaffolds, and language supports (Santos, Darling-Hammond & Cheuk, 2012). In order to effectively teach culturally relevant NGSS-aligned science, elementary teachers will need a strong grasp of the science content and how students learn it, learning progressions across the three dimensions, and using phenomena to drive unit design.

Despite the clear need for teacher knowledge of content, language development, and how students learn, research shows that many teachers struggle in all of these areas. Adamson, Santau & Lee (2013) followed a longitudinal study providing three years of professional development to 3rd, 4th and 5th grade teachers at schools within a large urban district which had high levels of ELLs and students receiving free and reduced lunch. The authors aimed to shed light on teachers' self-reported strategies in both teaching science concepts, and teaching English language and literacy during science instruction. Special attention was paid to the differences reported by 3rd, 4th and 5th grade teachers because only 5th grade teachers were held accountable by the state for science instruction. 15 schools initially volunteered to participate, and of those, 6 completed all five years of the program. Schools were separated into treatment and control groups. 198 3rd-5th grade teachers participated.

The professional development program consisted of curriculum units for

students and teachers, and teacher workshops during the summer and throughout the

school year (Adamson, Santau & Lee, 2013). Units focused on the National Science

Education Standards (the precursor to the NGSS) and included teacher guides, student

books, trade books and materials. The units were designed to progress students

towards student-initiated scientific inquiry. Materials were translated into Spanish and

Haitian-Creole, which were the two dominant home languages of the student

population. Teacher guides included possible student misconceptions and how to

handle them, possible difficulties in teaching the materials with suggested strategies,

background science, assessment tools, extensions and homework assignments.

Researchers observed teachers in the classroom and conducted post-

observation interviews (Adamson, Santau & Lee, 2013). The interviewers used specific

events from the lesson in their questions, but the overall structure of the interview

questions remained consistent from teacher to teacher. Interviews were coded in two

areas: strategies that supported science learning, and strategies that supported

language development. Additional codes emerged as the researchers began analyzing

the data. Inter-rater agreement was 90%. The entire data set was cross-validated to

ensure consistency.

After analyzing the data, researchers found that the most commonly reported

strategies used by teachers while teaching science were: connecting learning to prior

experience, identifying student difficulties during instruction, and engaging students in

hands-on learning (Adamson, Santau & Lee, 2013). The least common strategies were

planning and carrying out investigations, asking scientific questions and making predictions or hypotheses. The most frequently reported strategies for teaching language were using multiple modes of representation, using language in multiple contexts, and developing science vocabulary. Additionally, for teachers who participated in the full three year program, the data showed an increase in strategies including modifying instructional language for various levels of language proficiency, using language in multiple contexts, and allowing ELLs to discuss science in their home language.

The researchers concluded that they may not have been explicit enough with teachers about structuring learning so that students move on a continuum of inquiry towards planning their own investigations. They also found that even after three years in the program, teachers did not commonly report implementing the strategy of allowing students to discuss science in their home language. In future iterations, researchers recommend that professional developers provide classroom support, such as co-teaching and modeling.

Other researchers have raised the question: does PCK describe all aspects of teaching that are important (Turkan et al., 2014)? From this question arose the construct of Content Knowledge for Teaching, or CKT, which argues that teachers pull from their content knowledge for teaching in order to engage their students in the tasks at hand (Ball, Thames, & Phelps, 2008). However, CKT does not necessarily take into account the unique knowledge needed to teach special populations such as ELs. Turkan et al. (2014) propose Disciplinary Linguistic Knowledge, or DLK, as an important

construct for teaching academic content to English Language Learners. Their paper

"Proposing a Knowledge Base for Teaching Academic Content to English Language

Learners: Disciplinary Linguistic Knowledge" synthesize previous research and thinking

in both systemic functional linguistics (SFL) and academic language to describe DLK.

After analyzing several meta-analyses on academic language, Turkan et al. (2014)

describe academic language as the language used within a discipline that is distinct from

everyday language. Understanding the difference between everyday language and

academic language allows educators to identify the language demands inherent to

academic language (Bailey & Butler, 2003). A more recent meta-analysis on academic

language conducted by Anstrom et al. (2010) concluded that effective teachers

understand academic language beyond teaching students academic vocabulary, but

rather provide opportunities for students to authentically use the language of the

discipline.

Systemic functional linguistics (SFL) is a theory that focuses on the social

component of language (Turkan et al., 2014). While there is crossover between

academic language and SFL, SFL is not only concerned with the language register of a

given discipline, but the awareness of how to use language to express ideas and make

meaning within an academic context. This disciplinary discourse requires that

participants not only understand the academic language of the discipline, but also what

it means to be a part of the community of that discipline.

Based on the convergence of research around academic language and SFL,

Disciplinary Linguistic Knowledge (DLK) is a term used to describe the language

awareness teachers need when teaching content to English learners (Turkan et al., 2014). DLK is defined as "the linguistic knowledge base that all teachers of ELLs need to facilitate students' understanding of oral and written discourse within a discipline and their use of language in ways that allow them to actively participate in the disciplinary discourse" (p. 9). To do this, teachers must be able to first identify the "linguistic features and choices" (p. 9) of disciplinary discourse, and also have the ability to guide students in what those features look and sound like through modeling while making their own language choices explicit. Teachers also need to know the possible misconceptions that ELs might hold as they learn the language of each discipline.

Turkan et al. (2014) conclude by reiterating that Disciplinary Linguistic Knowledge be considered as an important knowledge base for any teachers working with English learners. Language must be taken into account when engaging students in reading, writing, listening, speaking and thinking in the language of any discipline. Teachers need to first have a strong understanding of their own use of disciplinary language, so that they can then be explicit in modeling that language use for their English learner students. They also need to make the language of the discipline accessible for their students with limited English, by unpacking both the language demands within the content area, and the linguistic choices that students will need to make in order to express their ideas.

Best Practices for Teacher Professional Development

With any kind of educational reform, professional development models need to be designed in ways that use research-based practices to support teachers as they

23

shift their practice in the classroom. In a large-scale study, Penuel et al. (2007) studied

the results of 454 teachers who were participants in an Earth science program called

GLOBE, delivered by 28 professional development providers, in order to learn more

about what elements of the program led to successful outcomes in science inquiry for

teachers and students. Data sources included surveys given to professional developers

and teachers that captured information about the implementation of the program, as

well as objective data from student work that was collected by the GLOBE program.

Researchers wanted to know what specific aspects of GLOBE professional development

led to greater program implementation, what kinds of program activities correlated to

changed teacher practice, and the possible effects of follow-up after the conclusion of

the program.

The GLOBE program is an international educational program that involves

students in using measurement tools to take data about weather in their area, and then

to report that data to GLOBE's website (Penuel et al., 2007). The program has taken

steps to promote inquiry both in the student and professional development experience,

such as providing visualizations of student data in an effort to encourage students to ask

and answer their own questions about the data, and examples of student-led projects

that can be accessed through the website or at GLOBE conferences. Professional

development centers around regional hubs which provide relevant, localized

experiences. However, this also means that professional development experiences vary

depending on location and available funding. GLOBE's website contains an extensive

database of student-collected data, making it a particularly good program for study, as 24

student data reporting in the classroom can also be observed by researchers. This study examined results from teachers in different regions in an attempt to discover if links could be found between professional development experiences and teacher and student outcomes.

The surveys used to collect information from teachers and professional developers were validated through expert review. To capture the level of program implementation, three dependant variables were measured: data reporting, protocol use, and preparation for student inquiry. To capture any change in teacher outcomes, two dependent variables were identified: knowledge of pedagogy and teacher change in practice. Data reporting and protocol use were straightforward: researchers looked to the uploaded data on the GLOBE website. Preparation for student inquiry involved asking teachers about their practice in relation to student inquiry, including long-term student projects, hands-on and field based projects, and analyzing the resultant interview data in multiple ways. Knowledge of pedagogy was determined through asking teachers to reflect on their feelings about their own science knowledge, use of assessments and technology, and abilities in teaching diverse students. Changes in teacher practice were determined by asking teachers about their use of technology and assessments, as well as instructional methods.

Independent variables of the study included barriers to success, materials used, whether the professional development was reform based (e.g. professional learning communities, coaching) or traditional (e.g. workshop), grade level of teachers served, coherence of the program and level of collaboration. Follow-up after the end of the

program was also measured by asking teachers about additional support they received. All data was analyzed using descriptive statistics and correlational analysis.

Researchers found that aligning the GLOBE professional development with state standards, and focusing on implementation with existing curriculum, had a positive effect on teacher self-reported readiness to engage students in science inquiry (Penuel et al., 2007). Similarly, the GLOBE programs which were most closely aligned to local district goals and reforms showed greater protocol use by teachers. Results also showed that collective participation and reform-based experiences had a positive effect on teacher change. Follow-up teacher support with the technology needed for student reporting was found to positively correlate to student data collection and protocol use, as well as teacher knowledge.

When considering professional development needed for teachers working with English learners, Lindahl and Watkins (2015) suggest that professional developers move away from a toolkit model for professional development, where specific strategies and activities are promoted, and instead move towards fostering Teacher Language Awareness, or TLA. TLA is comprised of three domains: the user domain, which is the teacher's own awareness of how English is used by themselves and different kinds of English learners, the analyst domain, which is the teacher's understanding of language forms, functions and registers, and the teacher domain, which is the teacher's ability to plan instruction that is engaging and culturally and linguistically appropriate for English learners (this could also be referred to as PCK, see above). In order to focus on teacher TLA in the analyst domain, Lindahl and Watkins (2015) suggest that professional

developers provide teachers with opportunities to construct Language Objectives for students, and identify language demands inherent in those Language Objectives through the use of a tool or framework.

In an example of research that focused on professional development aimed to increase teachers' language awareness in science, researchers worked with teachers at 10 Texas elementary schools identified as having high EL populations (Carrejo & Reinhartz, 2014). Throughout a year-long professional development program, teachers were trained in using the 5E instructional model: Engage, Explore, Explain, Elaborate, Evaluate (Bybee et al., 2006). Throughout the year, teachers and researchers considered the role of language in each instructional phase. Emphasis was placed on "co-developing of language literacy and content literacy through contextualized instruction" (Carrejo & Reinhartz, 2014, p.345).

Carrejo and Reinhartz (2014) found that despite a perceived need to focus solely on language to move students forward in their language acquisition, language and content literacy could be developed simultaneously. Teacher interviews indicated that the hands-on nature of inquiry science supported students at all levels of English proficiency, and the use of science notebooks promoted writing in science, which supported students at all language proficiency levels. In addition to qualitative evidence that students were benefiting from integrating language and science content development, researchers found statistically significant gains in fifth grade standardized test scores in both science and reading for all schools where teachers participated in the professional development. Carrejo and Reinhartz conclude that the structure of

27

teacher professional development is less important than emphasizing the simultaneous acquisition of language and content.

In another study focusing on elementary school teachers in three California schools with a high percentage of English learners, researchers tested a "blended" curriculum that emphasized language and science learning objectives simultaneously (Zweip et al., 2011). In all, 60 teachers, three principals and six English Language Development (ELD) teachers participated. Teachers at these schools were already using the 5E unit design (Bybee et al., 2006) to plan science instruction, although they reported that science was not regularly taught prior to entering into this study. ELD instruction focused on grammatical forms and functions, with pre-prescribed sentence frames and vocabulary front-loading (teaching the vocabulary used within a lesson before the start of that lesson) reported by research participants as common practices.

Through an iterative field testing process, teachers and researchers landed on a lesson plan design which emphasized the science learning as the "structural center" of the lessons, and then integrated ELD into the lesson design (Zweip et al., 2011). They made the decision, after field testing, to only front-load vocabulary which students would need to engage in the lesson (for example, names of materials that they would be engaging with) and to wait to teach the vocabulary that was tied to conceptual understanding until after students had engaged with the concepts. Attention was also given in the lesson design template to the language forms and functions which would be emphasized in each lesson.

After the field testing period, teachers used the revised blended lesson plan to plan integrated science and ELD lessons. Researchers measured the results through teacher interviews with 29 teachers, observations in ten classrooms, and interviews with two of the three principals. Interviews were recorded and the information coded, then organized into sets of coherent ideas. These sets of ideas were further organized into "key insights" that allowed researchers to draw conclusions about the effects of the blended curriculum program.

The first key insight from the program was that by combining ELD and science, the importance of science instruction was greatly heightened at the participating schools (Zweip et al., 2011). Prior to the study, science had been de-emphasized as teachers and principals focused on raising student test scores, but this new focus on science and language integration provided a reason to promote science instruction. Additionally, students were very excited about the shift from ELD to science instruction. According to teachers at the school, science is seen as a privileged activity while ELD indicates a deficit in student ability. Shifting away from ELD and towards science also improved student engagement and behavior. This unexpected outcome was expressed in teacher and principal interviews, as well as observed by researchers.

The second key insight from the blended program was that students were using more oral language in the classroom. Teachers reported that students were much more excited and willing to speak in class. This trend was observed by school staff not only in the classroom, but throughout the school day.

The third key insight was that teachers found themselves less concerned with the labels being placed on students, and more focused on their own teaching pedagogy (Zweip et al., 2011). Teachers expressed a realization that rather than focusing on the language levels of their students as a predictor of how they would do in class, they needed to focus on the kinds of teaching strategies they used to elicit student responses. Teachers also expressed the realization that their students may have been held back from performing in science not by their lack of science understanding, but rather because of their language abilities.

Researchers conclude that for blended curriculum to be successful, teachers need to understand nuanced decisions, such as when to emphasize language within a lesson and when to emphasize science content (Zweip et al., 2011). These decisions may need to be made as the lesson unfolds. In this particular project, teachers were more insecure about their ability to teach science than about their ability to teach language, thanks in part to their school district's ongoing professional development in ELD. However, teachers in other contexts who receive less training in ELD may struggle both with teaching ELD and science content to their students, and may require a different kind of support to use a blended curriculum lesson template like the one used in this study. The researchers conclude, much like in the previously mentioned study, that it is indeed possible to teach language and content simultaneously, and there are benefits to doing so for students, teachers and schools. Furthermore, offering science to students who are designated as ELs promotes equity, as these students often may be more likely to be denied science instruction in favor of ELD and English language arts.

In summary, the separation that once existed between language and content learning is slowly being broken down, as research is showing that students and teachers benefit from engaging, relevant and content-rich opportunities to use language throughout the school day. Furthermore, the model of expecting language specialists to teach English is no longer feasible as the number of English learners increases in schools across the country. These issues have been recognized by the writers of the framework that guided the English Language Proficiency Standards (CSSO, 2012), who envisioned these standards as a tool to enhance and support content standards, rather than to be used in isolation.

In addition to these shifts in how language is taught, teachers are also being asked to shift the way they teach science with the adoption of the Next Generation Science Standards (NRC, 2011). Although the standards are too new to have a body of research outlining best practices in professional development, other projects have shown that teachers need long-term, contextual, collaborative professional development in order to shift towards more inquiry-based teaching. Therefore, in light of the adoption of the Next Generation Science Standards and the English Language Proficiency standards in the state of Oregon, professional development in science should also take language development into account. This project takes the first step in this direction by asking participants in an 8-week NGSS summer course to engage in a Task Analysis process. The Task Analysis process allows for collaborative reflection on the knowledge and skills in the domains of content, analytical skills and language that are required to perform an NGSS-aligned science task.

Research Methods

Overview

This study is a mixed methods, post-test, quasi-experimental design which aims to answer the following research questions: 1. How well does the ELPA21 Task Analysis tool help participants understand the language demands of a science task when used as part of an NGSS summer professional development course? 2. How effective is a worksample and scoring rubric protocol for measuring the impact of the Task Analysis process?

The treatment is the incorporation of the Task Analysis process (see Appendix B) into an 8-week, four-unit course through Portland State University. There are 3 different courses, each with two sections. The course titles are: Change Over Time, Energy and Matter and Systems and Interactions. Each of the 6 course sections each had a different instructor, with the exception of the two Change Over Time sections which were taught by the same instructor. The instructors, with the researcher, principal investigator from the MSP grant, and course coordinator, formed the Instructional Leadership Team. The Instructional Leadership Team communicated on an ongoing basis to decide on learning outcomes for the NGSS courses, as well as common learning experiences for the participants, that could be measured through the research. It was decided that the treatment for the participants in this study would include instruction on the use of a "Task Analysis" tool (see Appendix B) that guides users through a consideration of the language demands of a content-based task in the areas of content, analytical skills and language.

Table 1: Definitions of uses of "N"

N_1	Number of paired responses scored
N_2	Number of participants represented

Table 2: Notation of study design

	Week 1: June 2016	Week 2-7: July-August 2016	Week 8: August 2016	Post-course: August- December 2016
N_2	O^1	X	O^2, O^3	O^4, O^5

Key:
N_2: non-randomized group
O^1: Task Analysis pre-assessment
X: Treatment
O^2: Course observations
O^3: Task Analysis post-assessment
O^4: Participant interviews
O^5: Instructor focus group

Table 2 is a depiction of the research design of this study. Week one of the project occurred the week of June 27, 2016, and the final sessions of the courses occurred the week of August 15, 2016. Post-course participant interviews and the instructor focus group took place between August 19th through the end of December, 2016.

Table 3: Dependent variables of study, with corresponding instruments and sources

Variable Measured	Instrument	Type	Source
• Effect of the Task Analysis process on understanding of:	Task Analysis Post-Assessment	Short answer	ELPA21 Task Analysis

o Content domain
  o Analytical Skills domain
  o Language domain
	Task Analysis Rubric	Four point scale rubric	--

• Instructor understanding of the Task Analysis process • Instructor feelings about the value of the Task Analysis process	Instructor Focus Group	Focus group discussion	--
• Participant understanding of the Task Analysis process • Participant feelings about the value of the Task Analysis process	Participant Interviews	Semi-structured interviews	--

The Task Analysis tool was used as the basis for the pre-assessment that was administered on Week 1 of the courses. The pre-assessment measured a baseline understanding in the domains of language, analytical skills and content (see Table 3 and Appendix C). Over the following weeks of the course, participants engaged with the ELPA21 Professional Development online module focusing on the Task Analysis process (Module two, n.d.), and then engaged in the Task Analysis process in small collaborative groups as part of an in-person class. On week 8, participants took the post-assessment (Appendix D), which measured the dependent variable of change in participant

understanding in the domains of language, analytical skills and content. Qualitative data about the Task Analysis process was collected through a focus group with course instructors, and interviews of course participants. Pre and post-assessment short answer data was scored using the Task Analysis Rubric (see Appendix A), which was developed for this project.

Participants

The participants for this study came from a variety of public and private schools in the Portland Metro area, as well as the Oregon coast. Course participation came from two populations: teachers participating in the 3-year MSP program (cohort group), and any K-12 teachers at large who heard about the courses and chose to enroll (at large group). Of the at large teachers, one was an informal educator, one was a preschool teacher and one was a pre-service teacher. All cohort teachers had their course costs reimbursed by the MSP K-6 Instructional Specialist grant funds. Teachers from the at large population who worked in the two largest Portland metro area school districts were also reimbursed for the cost of the course. All teachers who participated in the summer course, whether they were part of the cohort or at large group, could opt in to being included in the research. This provided an opportunity to examine results for participants who may be brand new to the NGSS in the at-large courses with teachers who had already had deep, ongoing professional development in the NGSS.

Course participants who consented to participate in the research ($N_2=64$)

26.6%

73.4%
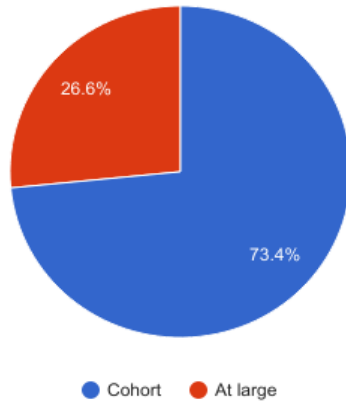
● Cohort   ● At large

Course participants who declined to participate in the research ($N_2=8$)

25%

75%
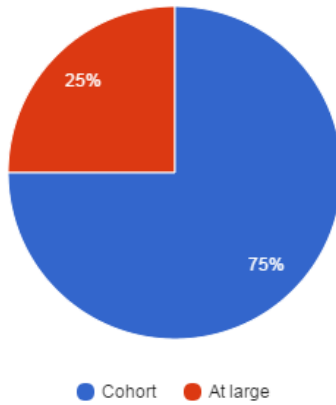
● Cohort   ● At large

Figure 2: Participation in the research - Cohort vs at large populations

In total, 87 educators registered for the NGSS summer courses, and 72 participants submitted the pre-assessment survey on the first day of the course. Of these, 64 consented to participate in the research, and eight declined. The Cohort and At large populations were almost equally represented in the group who consented to participate as those who did not (see Figure 2 above). Of the participants who consented, 16 did not complete the post-assessment in its entirety, and they were not

36

included in the research. 48 participants chose to participate in the study, *and*

completed the pre and post-assessment. Two participants were enrolled in two

different courses simultaneously, which meant that there were 50 completed pre and

post-assessments.

Table 4: Participant self-reported professional development history

	0	1-2	3-4	5+
Not including this course, how many science professional development sessions have you participated in within the last 12 months?	7	6	10	25
Of these sessions, how many focused on the NGSS?	8	8	8	24
Not including this course, how many English language learner-focused professional development sessions have you participated in within the last 12 months?	14	18	5	11
Of these, how many focused on the ELP Standards?	19	21	5	3

Participants reported the number of professional development sessions they

have attended in the last year as part of the pre-assessment. The results are presented

above in Table 4 above. Below, these results will be broken down into the Cohort and At

large participant populations for the sake of comparison.

Cohort (N₂=14)
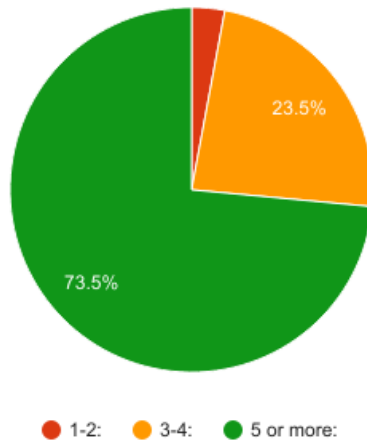
At large (N₂=34)
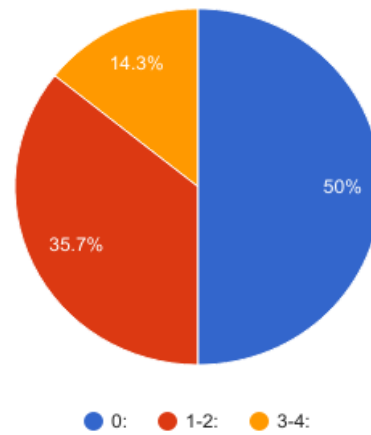
Figure 3: Not including this course, how many <u>science</u> professional development sessions have you participated in within the last 12 months?

Cohort (N₂=14)
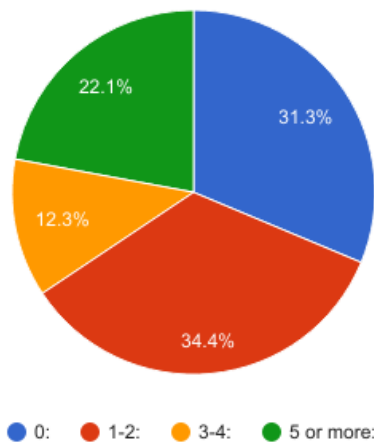
At large (N₂=34)
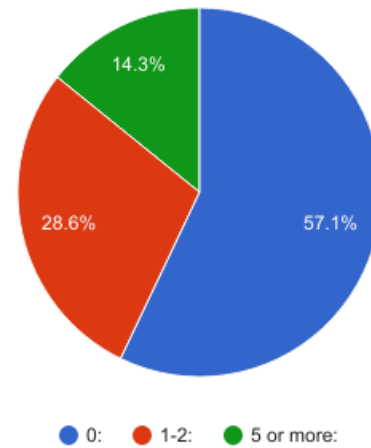
Figure 4: Not including this course, how many <u>English language learner</u> professional development sessions have you participated in within the last 12 months?

The participants from the cohort group had received more training on average in science (73% of participants reported attending 5 or more sessions in the last year) when compared to the at large group (no teachers reported attending this many sessions, see Figure 3 above). Cohort participants also reported receiving more training in teaching English learners than the at large group (see Figure 4 above). This may be partially explained by the MSP K-6 Instructional Specialist grant program's focus over the previous two years on providing guidance to participants on the role of language in science instruction. Participants also reported demographic information as part of their pre-assessment, which is detailed below in Table 5.

Table 5: Participant demographics

Gender	Male	Female		
	8	40		
Grade level[1]	**PK-2**	**3-5**	**6-8**	**9-12**
	15	26	7	1
School type	**Public**	**Private**	**Other**[2]	
	40	6	2	
Type of participation	**Cohort**	**At Large**		
	34	14		
Works with ELs?	**Yes**	**No**		
	37	11		
Current role[3][4]	**Classroom Teacher**	**ESL Teacher**	**Teacher on Special Assignment (TOSA)**	
	42	3	2	

[1]One K-8 teacher, one 2nd & 5th teacher, one K-5 teacher, one TOSA, two participants not currently working at a school
[2]Not currently working at a school
[3]Other roles: pre-service teacher (1), informal educator (1)
[4]One participant is both a classroom teacher and TOSA

Table 6: NGSS Summer Courses (by Instructor)

Course Title	Course Instructor	Number of participants
Energy and Matter (At large and Cohort)	1	9
Change Over Time (At large and cohort)	2	19
Systems and Interactions (Cohort)	3	2
Energy and Matter (Cohort)	4	10
Systems and Interactions (At large)	5	9

The 48 participants were broken into three different courses with six different sessions taught by five course instructors (one instructor taught two sessions of the Change Over Time course, see Table 6 above). Two of these six courses were titled Energy and Matter, two were titled Change Over Time and two were titled Systems and Interactions. Within the course sessions that shared the same title, one was designed for the cohort group of participants, and one was designed for at-large (although there was some participant crossover to accommodate needs in location and scheduling). Between the cohort and at-large sessions of the three courses, the curriculum was very similar although not completely identical. For the two sessions of each course, the tasks that the pre and post-assessments were based on were the same, and the pre and post-assessments were the same. Participants selected the courses based on interest or need: cohort participants from the three-year MSP Instructional Specialist grant were

required to take all three courses over the three years, so they selected their course based on which they had not yet completed. The at-large participants self-selected based on interest.

Treatment/Intervention

The PSU NGSS Summer Courses were initially designed to focus on improving teacher understanding of science and engineering content as outlined in the Framework for K-12 Science Education disciplinary core ideas. The courses are designed as "hybrid" courses, meaning they take place both in person and online. The three courses are organized by Crosscutting Concepts: Energy and Matter, Change Over Time, and Systems and Interactions. Cohort participants are required to take one course each summer as part of their commitment to the program, meaning by the end of the three-year grant, cohort teachers would have familiarized themselves more deeply with the Crosscutting Concepts, Science and Engineering Practices, and Disciplinary Core Ideas in Earth Science, Life Science, Physical Science, and Space Science. The bulk of each course was devoted to covering as much science content as possible, owing to the fact that many elementary teachers struggle with science content knowledge. After the first iteration, the courses were redesigned to include a greater emphasis on the Science and Engineering Practices, with each course focusing deeply on two of the practices.

There was concern that the MSP grant was not meeting its goals of integrating learning around how to best serve English learners, so that teacher leaders were being trained in how to help all students access the NGSS. At this point, it was decided that online professional development modules from the English Language Proficiency

42

Assessment for the 21st Century (ELPA21) initiative, which focus on walking teachers at all grade levels through a Task Analysis process in order to better understand the content, analytical, and linguistic demands of an NGSS-aligned task, would be included in each of the courses (see Appendix B for a copy of the Task Analysis tool). These modules were selected because they focused on the implementation of the ELP Standards, and addressing Oregon's language development standards would meet the needs of participating teachers.

The next step in the process was for the Course Coordinator and me to familiarize ourselves with the Task Analysis process through the online professional development module. Once we built our own understanding of the Task Analysis process, it was decided that the online module would work well as part of an online week of the course. We met with the course instructors and asked them to engage with the online module, and consider how best to incorporate it as part of their course. We had a second meeting to debrief what was learned, and what questions remained, about the online modules and the Task Analysis process.

Prior to the beginning of the courses, the five course instructors collaborated to decide upon a task from each of their sessions to focus on for the Task Analysis. The instructors were asked to focus on "language-rich" tasks, identified as tasks which focused on the NGSS practices of modeling, explaining, arguing or communicating (Quinn, Lee & Valdés, 2012). The three tasks selected for Task Analysis can be found in Appendix C: Pre-Assessment.

Each course participant engaged with the task and took a pre-assessment on the first day of the course (see Appendix C). After the pre-assessment was administered, each course instructor included the ELPA21 online module as part of their course during an online week. They also assigned course participants to collaboratively complete a Task Analysis of a task they had performed as part of an in-person session of the course (see Appendix B for an example of the Task Analysis tool).

In a similar process to the pre-assessment, instructors collaborated to identify three language-rich tasks from their final course session that could be focused on for the post-assessment Task Analysis. Three course instructors chose to spend class time discussing the Task Analysis process again, before the post-assessment was administered, and two course instructors simply administered the post-assessment without discussion time.

Interviews were conducted with participants who volunteered to speak with me after the conclusion of the courses. I interviewed seven participants who represented a range of grade levels and school types. I used a list of questions to guide the conversation, although not every participant was asked every interview question, depending on time. Each interview was 30 minutes long, conducted over the phone, and annotated (see Appendix E for interview questions). Once interview data was collected, I replaced each participant's name with a number to protect their anonymity.

The instructor focus group took place in August, after the conclusion of the courses. I created a list of questions to guide the discussion, although we did not end up discussing each question (see Appendix F for focus group questions). I led the

discussion and annotated the comments. The discussion lasted for one hour. Each instructor's name was replaced with a number to protect their anonymity.

Once participant data was collected, a rubric was created to score the results of the pre and post-assessment short answer questions (see Appendix C and D). In order to define the domains of content, analytical skills and language, I turned to the Task Analysis training module (Professional Development, 2014), Framework for K-12 Science Education (National Research Council, 2011), the NGSS Appendix A - Conceptual Shifts (NGSS Lead States, 2013), Framework for English Language Proficiency Development Standards Corresponding to the Common Core State Standards and the Next Generation Science Standards (CSSO, 2012), and the ELPA21 standards (CSSO, 2014).

I created a four point rubric scale, with categories of: 1 - Limited understanding, 2 - Emerging understanding, 3 - Proficient understanding, and 4 - Exemplary understanding (to view the rubric in its entirety, see Appendix A). In order to refine the rubric, I sought feedback from Tim Blackburn, who was involved in creating and training teachers to use the Task Analysis tool. He was able to provide suggestions that clarified language and improved representation of the domains. I also sought feedback on face validity from fellow graduate students and professors.

Further refinement of the rubric came as a course instructor, the Course Coordinator and I began using the rubric to score responses. We discussed changes needed to clarify the rubric for use by different scorers, as well as changes needed to better score the participant responses that we received. The final step to validating the rubric as a measurement tool was to compare scores for inter-rater reliability. Once

inter-rater reliability was established satisfactorily (see Results section), I used the finalized version of the rubric to score all of the participant responses.

Instruments

This study used several instruments to measure any change in teacher understanding of the Task Analysis process, as a result of engaging in the process over an eight-week summer NGSS course. Below, each instrument is listed and described in the order that it was used during the research study.

Pre-Assessment

In order to measure how engaging in the ELPA21 online module (Module Two: Task Analysis Process, n.d.) and the Task Analysis process might have shaped teacher understanding of the content, analytical, and linguistic demands of an NGSS-aligned task, we administered a pre-assessment (see Appendix C) to collect data on participants' baseline understanding of the domains of language, analytical skills and content. Participants were asked to answer the following six questions about a task they had engaged in as part of that day's class:

In order to successfully complete this task,

- What did you need to know in terms of content?

- What did you need to be able to do in terms of content?

- What did you need to know in terms of analytical skills?

- What did you need to be able to do in terms of analytical skills?

- What did you need to know in terms of language?

- What did you need to be able to do in terms of language?

The pre-assessment also asked teachers for demographic information, such as whether they work with English learners in their school, how much professional development they have received in Science and Language prior to the summer course, and how confident they felt about teaching science to English learners.

Post-Assessment

The post-assessment asked the same six short answer questions about a new task that participants engaged in on the last day of the course (see Appendix D). I was able to observe each course while the post-assessment was administered, which provided information on the circumstances surrounding the post-assessment in each course, namely whether the instructor chose to engage participants in a whole-group discussion prior to the post-assessment being administered. This information was then used to further analyze the data.

Task Analysis Rubric

In order to quantify what growth may have occurred in teacher responses as a result of engaging in the Task Analysis process, I developed a rubric aligned to the domains of language, analytical skills, and content knowledge as defined by the ELP standards and the NGSS (see Appendix A). The rubric is based off of definitions within the frameworks that guided the NGSS and ELP standards, and the description of the domains as seen in the Task Analysis online module (NRC, 2011; CCSSO, 2012; Module Two: Task Analysis process, n.d.). Each participant response was evaluated on a four point scale, from 1 (emerging understanding) to 4 (exemplary understanding). Using the final iteration of the rubric, inter-rater reliability was established at .664 with the first 47

instructor, and .899 with the second (see Tables 18 and 19). An item analysis was performed, which showed statistically significant change from the pre to post-assessment in every question except for Question 4 (see Results section), as well as normal score distribution (see Appendix G).

Instructor Focus Group

After the conclusion of the courses, I met with the course instructors to discuss the Task Analysis process. I created a list of questions (see appendix F) based on information I was hoping to gain, as well as the goals of the grant. Focus group questions were shared with the MSP grant leadership team and evaluator, as well as professors at PSU in order to establish face validity.

Participant Interviews

After the conclusion of the courses, I interviewed course participants to discuss the Task Analysis process. Interviews took place from August to December, 2016. I created a list of interview questions based on what I as a researcher, as well as the Grant, wanted to know about the participants' experiences with the Task Analysis (see Appendix E). Questions were shared with the MSP grant leadership team and evaluator, as well as professors at PSU in order to establish face validity.

Data analysis

Once the pre- and post-assessments were collected, participant names were replaced with numbers to protect their anonymity. The results were scored on a scale of 1-4, with 1 being "Emerging" and 4 being "Exemplary." These scores were analyzed in three different ways: by the percent optimal scores, by the movement in rubric scores

(ex: movement from a 1 to a 2 was measured as +1), and using a paired *t* test to determine statistical significance and effect size. The percent optimal scores were calculated by counting the number of scored responses that fell in the 3 (proficient) or 4 (exemplary) range. The benefit of using this method was that it allowed me to determine which percentage of our participants were demonstrating an understanding of the Task Analysis without relying on averaging scores, a less desirable method for capturing qualitative results. However, calculating percent optimal did not always show the growth that occurred, because a participant could have scored a 3 on the pre-assessment and a 4 on the post, and both would be considered in percent optimal range. To capture that growth, I also calculated movement along the rubric scale as either positive movement, negative movement, or no movement. This gave me two ways to look at the change that may have occurred between the pre- and post-assessments.

The second treatment of the scores was to measure the movement of rubric scores from the pre to post-assessment. A movement of 1 would indicate that a participant scored one point higher on the post-assessment than the pre-assessment. A negative number would indicate that a participant scored a lower score on the post-assessment than the pre-assessment. This method did not capture how highly participant responses scored on the rubric, but did capture any positive or negative change in scores.

The third treatment of the scores was to establish statistical significance and effect size. Statistical significance was determined by running a paired *t* test on the pre

and post-assessment scores to determine that the treatment had an effect. A Cohen's *d* test was used to determine effect size of the observed changes.

Finally, responses in the content and language domains were reviewed for mentions of "vocabulary." Whenever the word "vocabulary" was observed in a response, it was counted, as well as instances when the entire response centered around discussing vocabulary. These tabulated responses were then reported as percentages of all responses received.

Results

The purpose of this project was to determine the effectiveness of using the Task Analysis tool as a way to promote teacher understanding of the shifts in thinking around learning that are required by the NGSS and ELPA21 standards. In keeping with the spirit of the NGSS and ELPA21 standards, performance-based pre and post-assessments were used, and a rubric was created to quantify the responses. In addition to the pre and post-assessments, teacher participants were interviewed, and instructors participated in a focus group to discuss the experience of incorporating the Task Analysis into their courses.

Participant scores showed statistically significant growth across the three domains from the pre- to post-assessment, with one exception that will be discussed later. More insight into the experience was gained from the interview process, as well as the instructor focus group and course observations.

Participant self-reported readiness
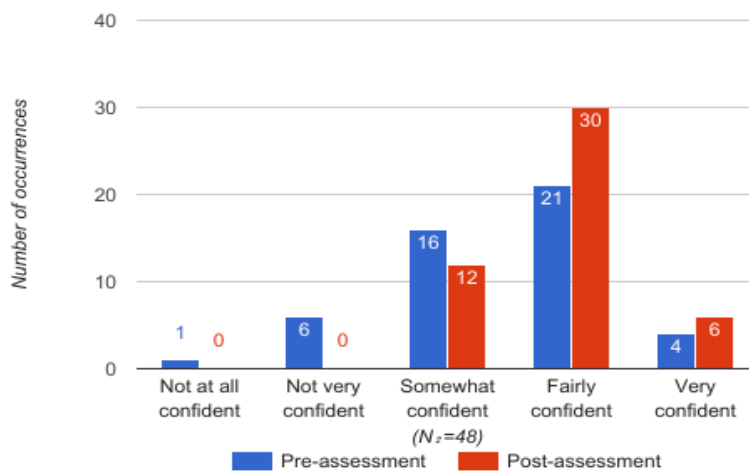
Figure 5: How confident do you feel about teaching NGSS-aligned science to English learners?

51

Table 7: Paired samples test - How confident do you feel about teaching NGSS-aligned science to English language learners?

| | Paired Differences | | | | | | | |
| | | | | 95% Confidence Interval of the Difference | | | | Sig. (2-tailed) |
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	
Confidence Pre - Confidence Post	-.43750	.82272	.11875	-.67639	-.19861	-3.684	47	.001

Participants were asked to self-evaluate their readiness for teaching NGSS-aligned science to English learners on a Likert scale, from Not at all confident to Very confident (see Appendices C and D). Change from the pre to post-assessment can be observed in the histogram above (see Figure 4). A two tailed *t* test showed that there was a significant difference in the scores for the pre-assessment (M=3.44, SD=0.9) and post-assessment (M=3.88, SD=0.61); t(47)=3.68, p = 0.0006. Cohen's *d* showed an effect size of 0.5723, showing a medium effect.

Content Knowledge
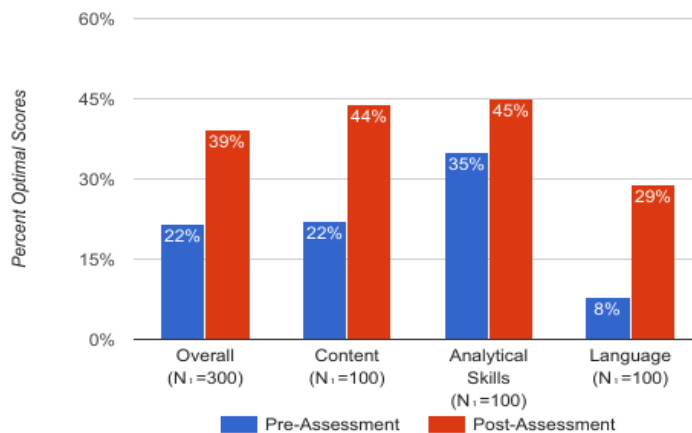
Figure 6: Percent optimal scores overall and by domain

The results from all three domains - content knowledge, analytical skills and language - were analyzed using four statistical measures: percent optimal scores, movement of pre- to post-assessment scores, a two-tailed paired *t* test, and a Cohen's *d* test. The purpose of measuring percent optimal scores was to determine the percentage of scores that fell in the optimal range - defined as a 3 (proficient) or 4 (exemplary) on the four-point rubric scale. When asked what they needed to know and be able to do in terms of content knowledge, participant scores rose from 22% to 44% in the percent optimal range (see Figure 1). These results showed growth, while also negating the possibility of a ceiling effect (where too many responses are scored at a 4 on the pre-survey to make the rubric a useful tool). The post-assessment scores of 44% in the percent optimal range suggest that future uses of the Task Analysis and rubric could have the potential to show yet even more growth.
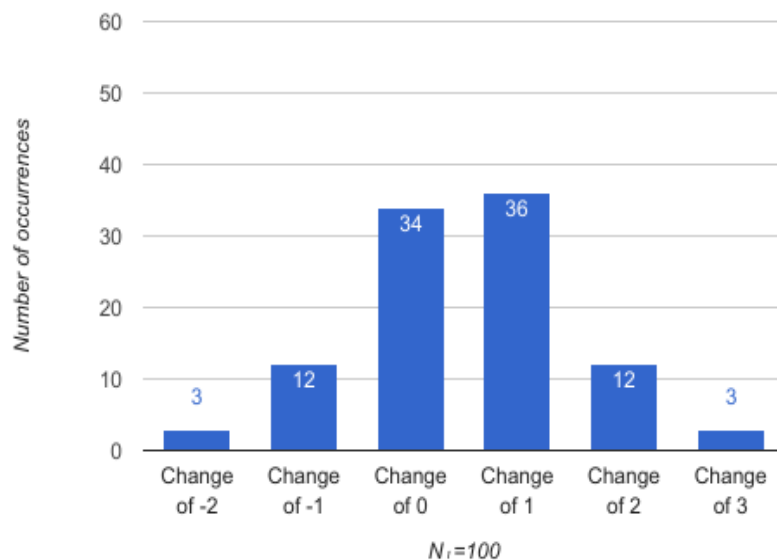
Figure 7: Pre to post score change in content

The next analysis involved looking at change from the pre to post-assessment scores, as shown in Figure 7. Positive numbers indicate positive growth (ex: moving from a 3 to a 2 would be recorded as a change of -1). In the domain of content, growth was observed in 51 out of 100 responses, no change in 34 responses, and negative change in 15 responses. This measure indicates that for approximately half of the responses received in the domain of content knowledge, scores from the post-assessment showed growth.

Table 8: Paired samples test - content

	Paired Differences							
				95% Confidence Interval of the Difference				Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	
Content Pre - Content Post	-.51000	1.04924	.10492	-.71819	-.30181	-4.861	99	.000

A paired *t* test was used to compare the results with the probability of a null hypothesis: that using the Task Analysis tool over the course had no effect on teacher response scores in the content domain. There was a significant difference in the scores for the pre-assessment (M=1.92, SD=0.72) and post-assessment (M=2.43, SD=0.79); $t(99)=4.86$, $p = 0.0001$. This allowed me to reject the null hypothesis, instead indicating that use of the Task Analysis tool over the summer course had an effect on teacher response scores in the content domain. Cohen's *d* was 0.675, showing a medium effect size (Sullivan & Fein, 2012).

Analytical Skills
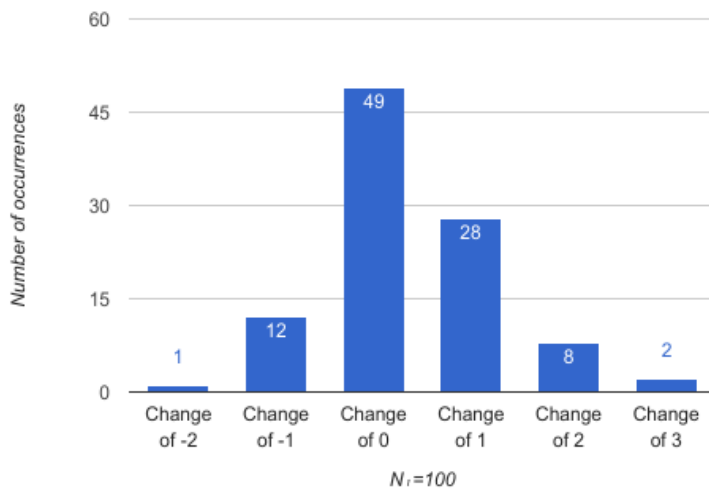
Figure 8: Pre to post score change in analytical skills

When asked what they needed to know and be able to do in the domain of

analytical skills to successfully complete a task, participant scores rose from 35% to 45%

in the percent optimal range (see Figure 6 above). Growth was observed, but this was

the domain that showed the least growth of the three. Also noteworthy is the fact that

percent optimal responses in the Analytical Skills domain were higher in the pre-

assessment than the other two domains, indicating that teachers entered into the

course with a stronger understanding of Analytical Skills (as defined by the Task Analysis

process) than language and content.

In the second measure, movement from pre to post-assessment responses, we

observed positive change in 38 of 100 responses, no change in 49 responses and

negative change in 13 responses (see Figure 8). While growth was observed, this domain

indicated the weakest positive change of the three.

55

Table 9: Paired samples test – analytical skills

| | Paired Differences | | | | | | | Sig. (2-tailed) |
| | Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference | | t | df | |
				Lower	Upper			
Analytical Pre - Analytical Post	-.26000	.84829	.08483	-.42832	-.09168	-3.065	99	.003

A paired *t* test was used to compare the results with the probability of a null hypothesis: that using the Task Analysis tool over the course had no effect on teacher response scores in the analytical skills domain (see Table 9). There was a significant difference in the scores for the pre-assessment (M=2.24, SD=0.67) and post-assessment (M=2.50, SD=0.75); t(99)=3.065, p = 0.0028. This allowed me to reject the null hypothesis, instead indicating that use of the Task Analysis tool over the summer course had an effect on teacher response scores in the analytical skills domain. Cohen's *d* was 0.366, showing a small effect size (Sullivan & Fein, 2012).

Language
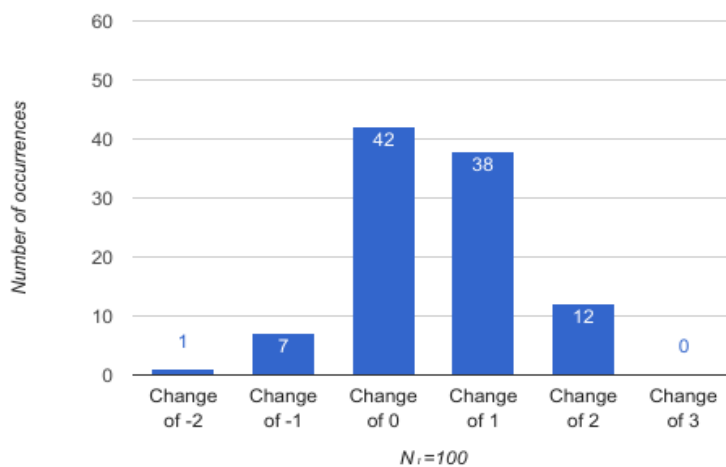
Figure 9: Pre to post score change in language

56

When asked what they needed to know and be able to do in the domain of language, participant scores rose from 8% to 29% in the percent optimal range (see Figure 6 above). Language was the domain with the lowest pre and post-assessment scores, but also showed more growth than analytical skills and was comparable to the content domain.

When analyzing the data by looking at the change in scores, growth was also observed. Out of 100 scored items, negative change was observed in 8 instances, no change was observed in 42, and positive change was observed in 50 (see Figure 9). This measure showed stronger growth than analytical skills, and comparable growth to the content domain.

Table 10: Paired samples test - language

	Paired Differences							
		Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
	Mean			Lower	Upper			
Language Pre – Language Post	-.53000	.83430	.08343	-.69554	-.36446	-6.353	99	.000

A paired *t* test was used to compare the results with the probability of a null hypothesis: that using the Task Analysis tool over the course had no effect on teacher response scores in the language domain. There was a significant difference in the scores for the pre-assessment (M=1.83, SD=0.63) and post-assessment (M=2.18, SD=0.64); t(99)=6.353, p = 0.0001 (see Table 10). This allowed me to reject the null hypothesis, instead indicating that use of the Task Analysis tool over the summer course had an effect on teacher response scores in the language domain. Cohen's *d* was 0.551,

showing a medium effect size (Sullivan & Fein, 2012).

Course comparison
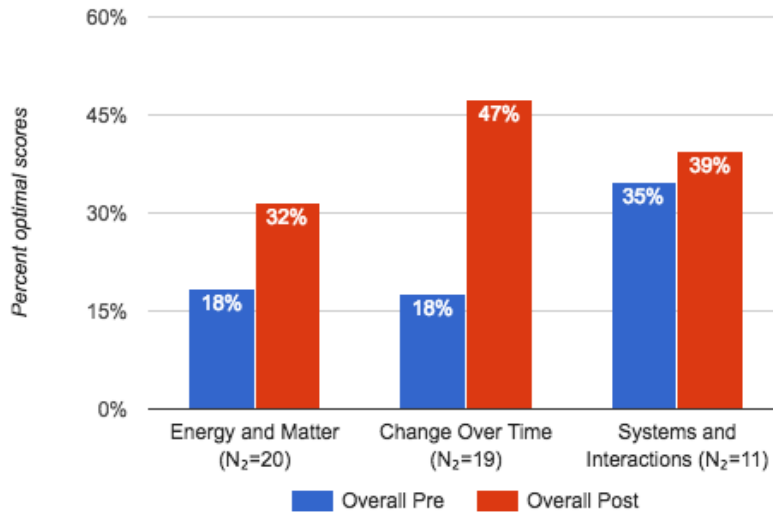
Figure 10: Percent optimal scores by course - overall

There were notable differences in the observed results in the three courses. Of the three, Change Over Time participants showed the most growth overall, with a change of 19% in percent optimal scores from the pre to post-assessment. Energy and Matter participants showed 14% growth, and Systems and Interactions participants showed 4% growth (see Figure 10).
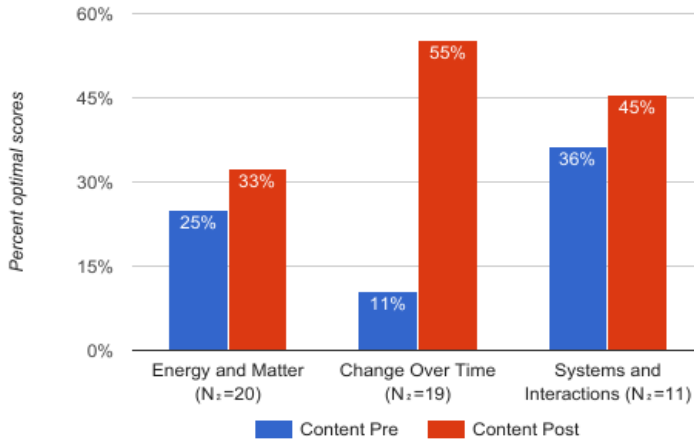
Figure 11: Percent optimal scores by course - content

Table 11: Paired samples test – course comparison in content

	Paired Differences							Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	
				Lower	Upper			
Pair 1 Energy and Matter Pre - Energy and Matter Post	-.45000	1.03651	.16389	-.78149	-.11851	-2.746	39	.009
Pair 2 Change Over Time Pre - Change Over Time Post	-.73684	1.03151	.16733	-1.07589	-.39779	-4.403	37	.000
Pair 3 Systems and Interactions Pre - Systems and Interactions Post	-.22727	1.06600	.22727	-.69991	.24537	-1.000	21	.329

When broken down further into the three domains, Change Over Time

participants showed the most growth in Content, with 44% more scores falling in the

percent optimal range (see Figure 11). There was a significant difference in the scores

for the pre-assessment (M=1.71, SD=0.65) and post-assessment (M=2.45, SD=0.83);

t=4.4035, p < 0.0001 (see Table 11). Cohen's *d* was 0.993, showing a large effect size

(Sullivan & Fein, 2012). Energy and Matter participants showed 8% growth in content

scores in the percent optimal range (see Figure 11). A paired *t* test showed there was a

significant difference in the scores for the pre-assessment (M=1.93, SD=0.76) and post-

assessment (M=2.38, SD=0.81); t=2.7438, p = 0.0091 (see Table 11). Cohen's *d* was

0.573, showing a medium effect size (Sullivan & Fein, 2012). Systems and Interactions

participants showed 9% growth in content scores in the percent optimal range (see

Figure 11). A paired *t* test showed there was not a significant difference in the scores for

the pre-assessment (M=2.27, SD=0.63) and post-assessment (M=2.5, SD=0.74); t=1.0, p

= 0.3287 (see Table 11). Cohen's *d* was 0.334, showing a small effect size (Sullivan &

Fein, 2012).

Figure 12: Percent optimal scores by course - analytical skills

Table 12: Paired samples test - course comparison in analytical skills

| | Paired Differences | | | | | | | |
| | | | | 95% Confidence Interval of the Difference | | | | Sig. (2-tailed) |
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	
Pair 1 Energy and Matter Pre - Energy and Matter Post	-.27500	.64001	.10119	-.47969	-.07031	-2.718	39	.010
Pair 2 Change Over Time Pre - Change Over Time Post	-.23684	.99822	.16193	-.56495	.09126	-1.463	37	.152
Pair 3 Systems and Interactions Pre - Systems and Interactions Post	-.27273	.93513	.19937	-.68734	.14188	-1.368	21	.186

Change Over Time participants showed a 12% growth scores in the percent optimal range in the domain of analytical skills (see Figure 12). There was not a significant difference in the scores for the pre-assessment (M=2.26, SD=0.76) and post-assessment (M=2.5, SD=0.76); t=1.4626, p = 0.152 (see Table 12). Cohen's *d* was 0.316, showing a small effect size (Sullivan & Fein, 2012). Energy and Matter participants showed 11% growth in analytical skills scores in the percent optimal range (see Figure 12). A paired *t* test showed there was a significant difference in the scores for the pre-assessment (M=2.2, SD=0.56) and post-assessment (M=2.48, SD=0.64); t=2.718, p = 0.0098 (see Table 12). Cohen's *d* was 0.466, showing a small effect size (Sullivan & Fein, 2012). Systems and Interactions participants showed 4% growth in analytical skills scores in the percent optimal range (see Figure 12). A paired *t* test showed there was not a significant difference in the scores for the pre-assessment (M=2.27, SD=0.7) and post-assessment (M=2.55, SD=0.91); t=1.3679, p = 0.1858 (see Table 12). Cohen's *d* was 0.345, showing a small effect size (Sullivan & Fein, 2012).
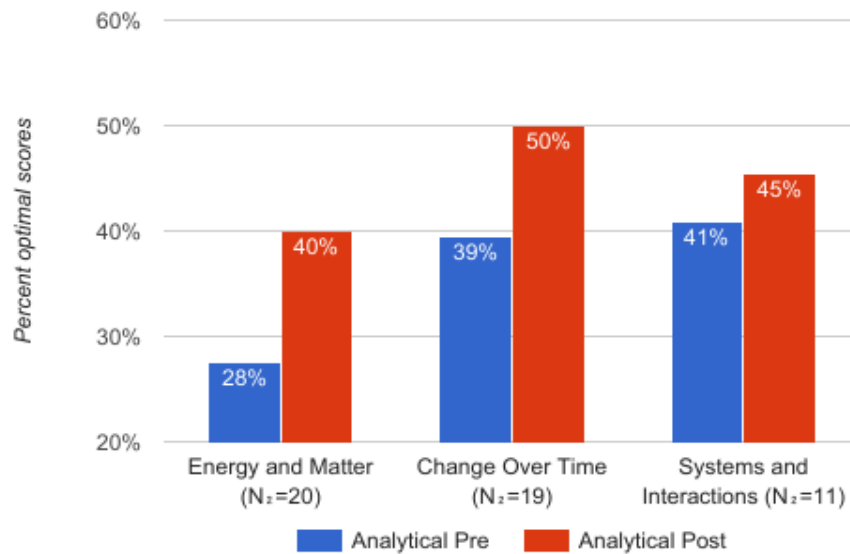
61
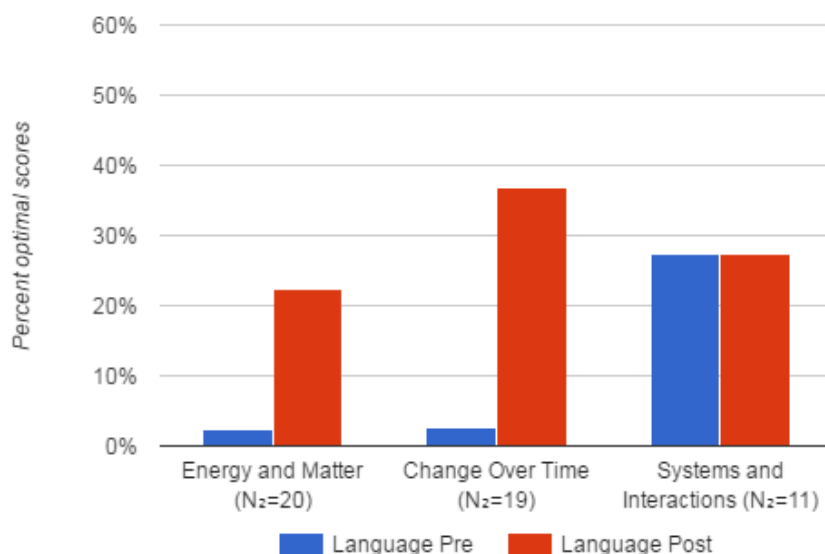
Figure 13: Percent optimal scores by course - language

Table 13: Paired samples test - course comparison in language

| | Paired Differences | | | | | | | |
| | Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference | | t | df | Sig. (2-tailed) |
				Lower	Upper			
Pair 1 Energy and Matter Pre - Energy and Matter Post	-.60000	.70892	.11209	-.82672	-.37328	-5.353	39	.000
Pair 2 Change Over Time Pre - Change Over Time Post	-.65789	.84714	.13742	-.93634	-.37945	-4.787	37	.000
Pair 3 Systems and Interactions Pre - Systems and Interactions Post	-.18182	.95799	.20424	-.60657	.24293	-.890	21	.383

In the domain of language, Change Over Time participants showed a 34% growth of scores in the percent optimal range (see Figure 13). A paired *t* test showed there was a significant difference in the scores for the pre-assessment (M=1.61, SD=0.55) and post-assessment (M=2.26, SD=0.64); t=4.787, p < 0.0001 (see Table 13). Cohen's *d* was 1.089, showing a large effect size (Sullivan & Fein, 2012). Energy and

Matter participants showed 20% growth in language scores in the percent optimal range

(see Figure 13). A paired *t* test showed there was a significant difference in the scores

for the pre-assessment (M=1.55, SD=0.55) and post-assessment (M=2.15, SD=0.53);

t=5.3529, p < 0.0001 (see Table 13). Cohen's *d* was 1.111, showing a large effect size

(Sullivan & Fein, 2012). Systems and Interactions participants showed no growth in

language scores in the percent optimal range (see Figure 13). A paired *t* test showed

there was not a significant difference in the scores for the pre-assessment (M=1.91,

SD=0.81) and post-assessment (M=2.09, SD=0.81); t=0.8902, p = 0.3834 (see Table 13).

Cohen's *d* was 0.222, showing a small effect size (Sullivan & Fein, 2012).

Time for discussion
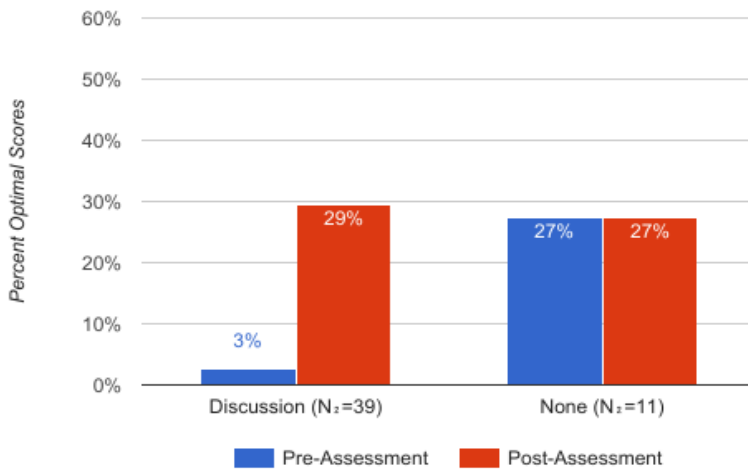
Figure 14: Percent optimal scores in language - discussion vs none

Of the 6 course sessions, 3 instructors chose to provide time for participants to discuss the Task Analysis before completing the post-assessment, and 2 chose to give the post-assessment without discussion. The discussions happened in the Change Over Time and Energy and Matter courses, while the Systems and Interactions courses did not engage in discussion prior to the post-assessment. When looking at the results from these courses, there was consistently more positive growth measured in the groups which had discussion time compared to those who did not. In the content domain, there was a growth of 26% for the group who had the opportunity to discuss the Task Analysis before taking the post-assessment, versus a 9% growth for the group who did not have a discussion (see Figure 14). In the analytical skills domain, there was a growth of 12% percent optimal scores for the discussion group versus 4% for the non-discussion group. In the language domain, there was a growth of 26% percent optimal scores in the

discussion group, versus no growth shown in the group which did not have a chance to

discuss.

Table 14: Paired samples test - discussion comparison in language

| | | Paired Differences | | | | | | |
| | Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference | | t | df | Sig. (2-tailed) |
				Lower	Upper			
Pair 1 Discussion Pre - Discussion Post	-.62821	.77475	.08772	-.80288	-.45353	-7.161	77	.000
Pair 2 No Discussion Pre - No Discussion Post	-.18182	.95799	.20424	-.60657	.24293	-.890	21	.383

A particularly strong signal was observed in the differences in language scores

between the groups who had time to discuss and the groups who did not (see Table 14).

A paired *t* test showed, for groups who has the opportunity for discussion, there was a

significant difference in the scores for the pre-assessment (M=1.58, SD=0.55) and post-

assessment (M=2.21, SD=0.59); t=7.1612, p<0.0001. Cohen's *d* was 1.105, showing a

large effect size (Sullivan & Fein, 2012). By comparison, for the groups who did not

engage in discussion prior to the post-assessment, there was no significant difference in

the scores for the pre-assessment (M=1.91, SD=0.81) and post-assessment (M=2.09,

SD=0.81); t=0.8902, p=0.3834. Cohen's *d* was .222, showing a small effect size (Sullivan

& Fein, 2012).

Grade level
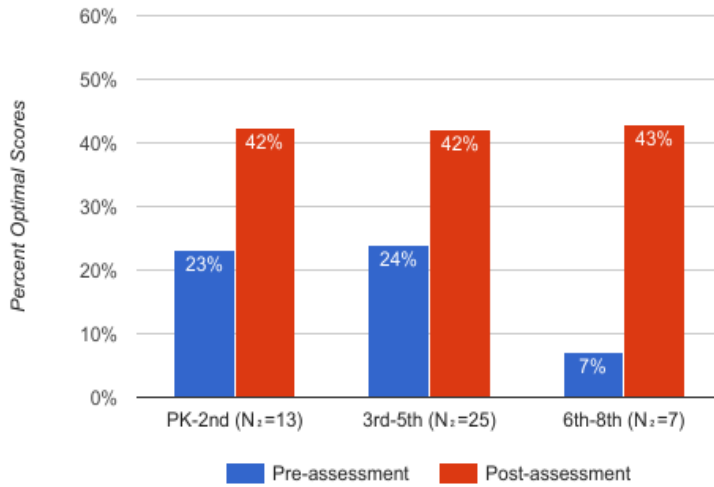
Figure 15: Percent optimal scores by grade level - content
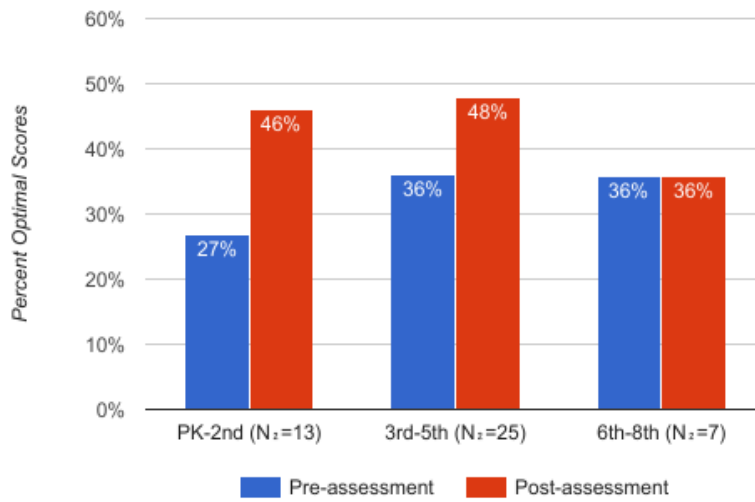
Figure 16: Percent optimal scores by grade level - analytical skills

66
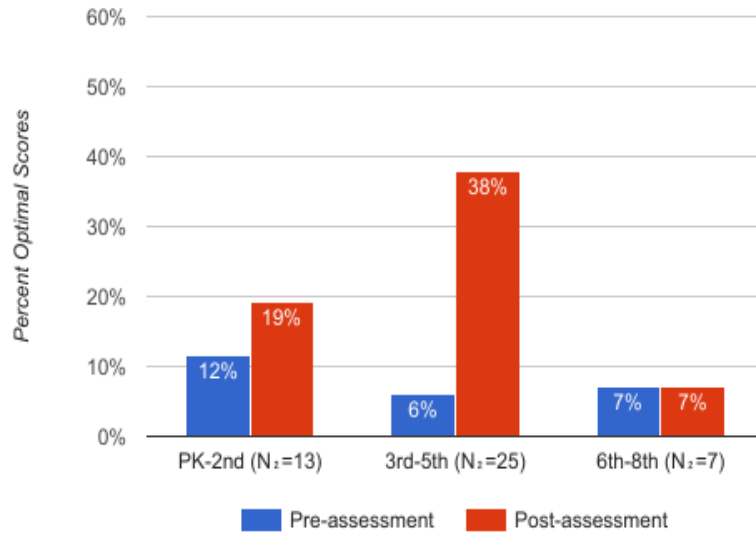
Figure 17: Percent optimal scores by grade level - language

When data was broken down by what grade level the participants are currently teaching, 6th-8th grade teachers did not show any growth in percent optimal scores within the domains of analytical skills and language, but showed the largest growth of any group in their understanding of content, with a change in percent optimal scores of 37%, compared to 19% for PK-2 teachers and 18% for 3-5th teachers. PK-2nd grade teachers showed the largest growth in the analytical skills domain, with a change in percent optimal scores of 19%, compared to 12% for 3rd-5th grade teachers and no change for 6th-8th grade teachers. 3rd-5th grade teachers showed the largest growth in the language domain, with a change of 32% percent optimal scores, compared to 7% for PK-2 teachers and no change for 6-8 teachers.

Teachers who currently work with English learners
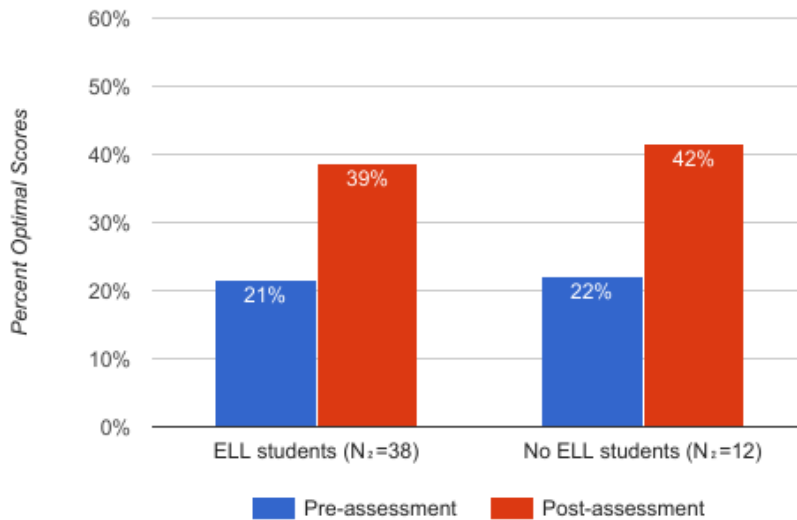
Figure 18: Percent optimal scores overall - English learner students vs none
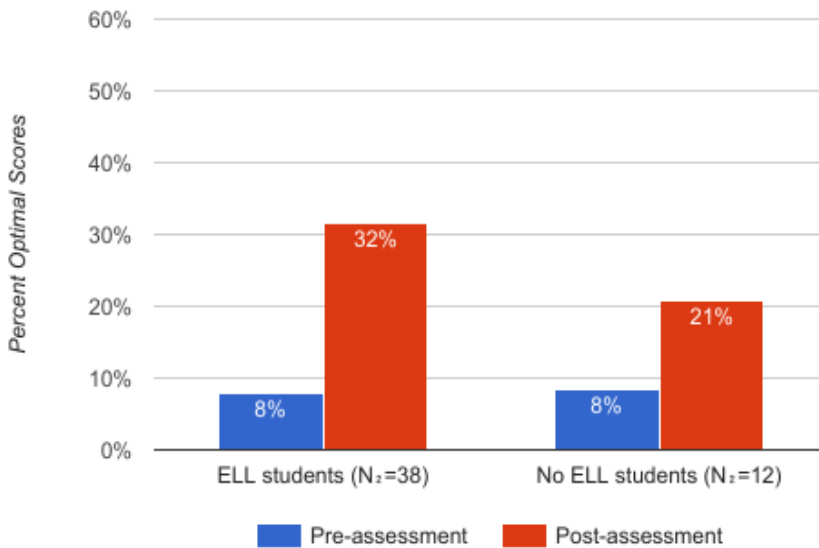
Figure 19: Percent optimal scores in language - English learner students vs none

Table 15: Paired samples test – English learner comparison in language

| | Paired Differences | | | | | | | |
| | Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference | | t | df | Sig. (2-tailed) |
				Lower	Upper			
Pair 1 ELs Pre - ELs Post	-.53947	.82366	.09448	-.72769	-.35126	-5.710	75	.000
Pair 2 No ELs Pre - No ELs Post	-.50000	.88465	.18058	-.87356	-.12644	-2.769	23	.011

Teachers showed comparable growth overall, whether they currently work with English Learners (ELs) in their classrooms or not: teachers who work with ELs showed an 18% growth in percent optimal scores overall, versus 20% growth for teachers who don't work with ELs (see Figure 18). When broken down by domain, there was a noticeable difference in the two groups' growth in the language domain, with teachers of ELs showing a growth of 24% versus 13% for teachers who do not work with ELs (see Figure 19). However, statistical analysis showed that these groups may have performed more similarly than initial analysis suggested. A paired *t* test showed, for teachers of ELs, there was a significant difference in the language scores for the pre-assessment (M=1.66, SD=0.62) and post-assessment (M=2.2, SD=0.63); t=5.709, p<0.0001 (see Table 15). Cohen's *d* was 0.864, showing a large effect size (Sullivan & Fein, 2012). By comparison, teachers who do not work with ELs also showed significant difference in the scores for the pre-assessment (M=1.63, SD=0.65) and post-assessment (M=2.13, SD=0.68); t(23)=2.7689, p<0.0001. Cohen's *d* was .752, showing a medium effect size (Sullivan & Fein, 2012).

Cohort vs At Large

This project provided the opportunity to compare the results of two populations of teachers: those who took the NGSS summer courses as part of a three-year Math Science Partnership grant commitment (cohort), and those who registered for other reasons (at large). From the onset of the project, the question was raised: would there be a difference in results between the cohort teachers and at large teachers?
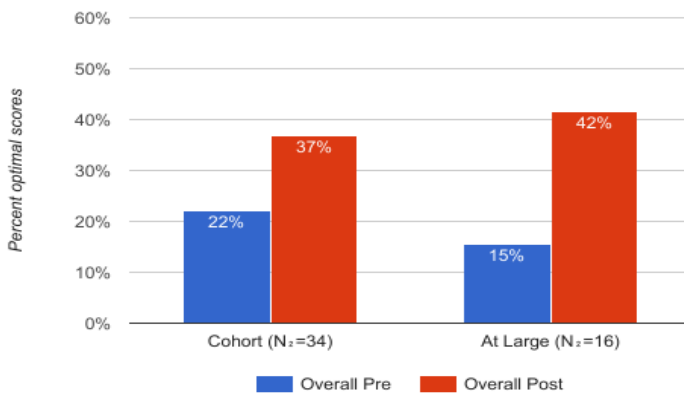
Figure 20: Cohort vs at large percent optimal scores - overall
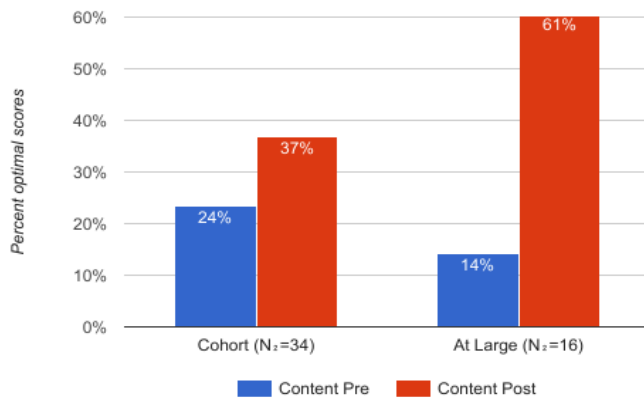
Figure 21: Cohort vs at large percent optimal scores - content

The results showed that there were some differences. Overall, when looking at the percent optimal scores of the two groups, the cohort group's scores changed from 22% to 37% in the percent optimal range, while the at large group's scores changed from 21% to 45% in the percent optimal range (see Figure 20). When broken down into the three domains, the difference in scores in the content domain is the most pronounced (see Figure 21). Cohort teachers came into the courses with a higher baseline of responses in the percent optimal range for content (24% for the cohort group as opposed to 19% for the at large group). However, the at large group scored 45% in the percent optimal range on the post-assessment, as opposed to 37% for the cohort group, showing that they outperformed their cohort colleagues.

Table 16: Paired samples test - cohort/at large comparison in content

	Paired Differences							
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
				Lower	Upper			
Pair 1 Cohort Pre - Cohort Post	-.42647	1.11055	.13467	-.69528	-.15766	-3.167	67	.002
Pair 2 At Large Pre – At Large Post	-.68750	.89578	.15835	-1.01046	-.36454	-4.342	31	.000

A paired *t* test showed, for the cohort group's content scores, there was a significant difference in the scores for the pre-assessment (M=1.91, SD=0.75) and post-assessment (M=2.34, SD=0.75); t(67)=3.1667, p=0.0023 (see Table 16). Cohen's *d* was 0.529, showing a medium effect size (Sullivan & Fein, 2012). By comparison, for the at large group, there was a significant difference in the scores for the pre-assessment

(M=1.94, SD=0.67) and post-assessment (M=2.63, SD=0.55); t(31)=4.3416, p=0.0001.

Cohen's *d* was 1.126, showing a large effect size (Sullivan & Fein, 2012).

Mentions of "Vocabulary" in participant responses

Initial reading of participant responses brought a trend to light that was later

confirmed by further counting and evaluating of responses: Teachers across the three

courses and five sessions emphasized vocabulary use and understanding across all three

courses when asked what students needed to know in the domains of content and

language.

Table 17: Percent of total responses that mentioned "Vocabulary" in pre-assessment

	What did you need to know in terms of content?	What did you need to be able to do in terms of content?	What did you need to know in terms of language?	What did you need to be able to do in terms of language?
Change Over Time (N_2=19)	32%	21%	89%	42%
Energy and Matter (N_2=20)	35%	10%	80%	35%
Systems and Interactions (N_2=11)	9%	0%	73%	27%

Table 18: Percent of total responses that mentioned "Vocabulary" in post-assessment

	What did you need to know in terms of content?	What did you need to be able to do in terms of content?	What did you need to know in terms of language?	What did you need to be able to do in terms of language?
Change Over Time (N_2=19)	37%	11%	84%	26%
Energy and Matter (N_2=20)	25%	15%	45%	15%
Systems and Interactions (N_2=11)	9%	0%	73%	36%

Many participant responses mentioned vocabulary as a part of their response. When asked "In order to successfully complete this task, what did you need to know in terms of content?," vocabulary was mentioned in a total of 14 out of 50 responses in the pre-assessment (see Table 17) and 13 out of 50 responses in the post-assessment (see Table 18). When asked the same question in the domain of language, 41 out of 50 responses mentioned vocabulary in the pre-assessment, and 33 out of 50 mentioned language in the post-assessment.

Table 19: Percent of total responses that <u>only</u> mentioned "Vocabulary" in pre-assessment

	What did you need to know in terms of content?	What did you need to be able to do in terms of content?	What did you need to know in terms of language?	What did you need to be able to do in terms of language?
Change Over Time **(N_2=19)**	26%	11%	58%	16%
Energy and Matter **(N_2=20)**	30%	0%	55%	20%
Systems and Interactions **(N_2=11)**	9%	0%	45%	9%

Table 20: Percent of total responses that <u>only</u> mentioned "Vocabulary" in post-assessment

	What did you need to know in terms of content?	What did you need to be able to do in terms of content?	What did you need to know in terms of language?	What did you need to be able to do in terms of language?
Change Over Time **(N_2=19)**	11%	0%	26%	5%
Energy and Matter **(N_2=20)**	5%	0%	10%	0%
Systems and Interactions **(N_2=11)**	0%	0%	64%	9%

Some participant responses *only* discussed vocabulary. When asked "In order to complete this task, what did you need to know in terms of content?" 12 out of 50 responses only discussed vocabulary in the pre-assessment (see Table 19) and 3 out of 50 responses in the post-assessment (see Table 20). When asked the same question in

the domain of language, 27 out of 50 responses only discussed vocabulary in the pre-assessment, and 14 out of 50 responses only discussed language in the post-assessment.

Interviews

While participant responses provided one view of the effects of using the Task Analysis tool during the summer courses, interviews provided another. I interviewed seven teachers who volunteered to speak with me. These teachers represented a range of grade levels taught, and came from a mix of school districts. My goals were to find out more about their experience with the task analysis, and whether they felt it would be a tool they would use again in the future, as well as more general comments and suggestions about the use of the Task Analysis within the courses. I kept the interviews to 30 minutes, and used a set of questions, although I did not ask every participant every interview question (see Appendix E for interview questions).

Of the seven teacher participants I interviewed, four indicated that they could see themselves using the tool, one had already used it with colleagues, one felt they would use it "down the road" when teachers were more ready and able to dive into the NGSS, and one expressed uncertainty over whether they truly understood how it was to be used: "I'm not sure if we even did [the Task Analysis] correctly so I wouldn't feel comfortable using it outside of the class. I'd want more familiarity with it." One participant who expressed confidence in using the Task Analysis at her site said: "We are colleagues and we can learn from each other. I think everybody would be up for it." Another teacher had already been introduced to using the Task Analysis, and had the

opportunity to use introduce it in a leadership capacity within her district. She described

her experience with using the Task Analysis:

> It was a district PLC Leadership Day, so I worked with all the schools in the
>
> district and a number of their leaders. In particular, what I'm trying to do is show
>
> good science instruction and that we're engaging in really great NGSS
>
> instruction. We did a task analysis...we went farther and discussed language
>
> scaffolds we could provide for students who aren't as proficient. It went really
>
> well. The only part we got caught up on was the analytical skills because people
>
> really wanted a list they could pick from when they just didn't understand what
>
> it was. I didn't really know how to give them a quick explanation.

Another participant touched on the collaborative nature of the Task Analysis process:

> I would probably not use [the Task Analysis] by myself, I would probably use it
>
> with colleagues. If we were doing PD looking at new science curriculum...we
>
> could use the tool to think about what is this not [doing] and what would we
>
> have to do? What would we have to pull in from other sources?

A few teachers discussed their increased awareness of the language that students were

being asked to use during their science tasks in the classroom:

> What the course did for me was to make me more aware of my need for
>
> integrating the ELL language demands into my lessons. I had some background
>
> knowledge already on the ELPs, so it helped me to learn how to integrate those
>
> language demands into my lessons better.
>
> If students do a PP presentation, I don't normally articulate the language

76

benefits or think about the language component of the lesson. [This was] a

chance to think more about those things and reflect.

I felt like I understood how to identify the language better. I understood how the

NGSS demanded the language but it helped me really look at an activity and

really pull out the language that was present. Is it hitting the NGSS? Is it

supporting the language? Or, how can I change this to bring out more language?

That was done really well, actually.

Focus Group

The focus group provided an opportunity for course instructors to discuss the

Task Analysis process as a group, and debrief how their courses went. Four of the five

course instructors were present, including the Course Coordinator, as well as the PI of

the grant and the grant evaluator. I led the discussion and annotated responses.

Two instructors shared that they had felt underprepared to lead the Task

Analysis, both due to their focus having been on conceptual understanding of science

and the NGSS, and a lack of time to prepare and receive professional development

around the process. Despite these challenges, one instructor did not view the addition

of the Task Analysis as a negative, but rather an opportunity: "For me, it was fascinating

to be a co-learner alongside the teachers and listen to their conversations to educate

myself and develop more understanding." Another instructor equated this challenge

with what they had been asking of teachers in regards to the science content: "With

77

NGSS we're trying to break down that scientists [are] over here, [the] rest of society [is] over here, we're trying to build language together."

All of the course instructors who participated indicated that they saw value in using the Task Analysis, and that their participants did as well. One instructor said, "The second time [that participants engaged in the Task Analysis] they made the comment that they could see it really does help to demonstrate and make clear the language demands required." Another instructor described the process: **"**How much analysis was going on - the content list was short but the analysis column was huge. Oh my gosh, the cognitive demands!" Elaborating on the process, a third instructor commented on the group work that characterized the Task Analysis, and how helpful it was: "First time I did it I grouped them by age groups - what age they teach. Second time [was] totally [the] opposite. This time it was great, they brought out the best in each other." Another instructor agreed: "Groups have really great conversations and help each other."

Throughout the conversation, some other interesting information emerged. The NGSS courses have evolved since their inception, when they were designed to be content-only courses. Since that time, course instructors have re-envisioned the courses to be more three-dimensional. As one instructor said, "To me that is a fundamental shift from thinking they need to develop their knowledge to [thinking] they need to be able to apply their knowledge in their classroom." The addition of the Task Analysis to the course was a new adjustment this year, and it felt disjointed to one course instructor.

She said:

> My own better understanding of [the Task Analysis] process would be very
> helpful. I see its value big time but I need some clarity of what the goals of the
> course are. It feels like it was thrown in there.

The Course Coordinator acknowledged that adding the Task Analysis to the NGSS
courses was an additional challenge for instructors:

> Time and payment in the spring was a limiting factor for all of you. What [we] did
> was not fair to ask of you. You're feeling like "I got this" and then your first day is
> totally planned out and you realize that you have to do the Task Analysis.

Despite these challenges, the instructor comments about their overall experience as a
course instructor were very positive. One said: "Being an instructor has vastly improved
my understanding of NGSS. This has been the most powerful science education
experience of my life." Another shared comments she was excited to hear from her
participants: "I can apply what I did last summer to this!" "Although I've done activities
before, all of the talking helps me understand better what scientists do."

Validation of the Task Analysis Rubric

The first step to validating the rubric as a measurement tool was to establish
face validity by sharing the rubric with colleagues and receiving feedback. The next step
was to compare scores for inter-rater reliability. Using my independently rated scores as
a comparison to the Course Coordinator's scores, I ran a Cohen's Kappa test. Our first
attempt yielded an agreement of .411 (Moderate Agreement). After editing the rubric
(including clarifying what a "proficient" score should include, as indicated above) and 79

further discussing our scoring decisions, we were able to raise our agreement to .664

(Substantial Agreement). Using scores generated from the refined version of the rubric, I

ran a Cohen's Kappa test on a course instructor's scores to compare them to mine, and

on the first attempt we had .799 (Substantial) agreement (Landis & Koch, 1977).

Table 21: Inter rater reliability with Instructor 4

Symmetric Measures

		Value	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance
Measure of Agreement	Kappa	**.799**	.065	9.537	.000
N of Valid Cases		60			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Table 22: Inter rater reliability with Instructor 5

Symmetric Measures

		Value	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance
Measure of Agreement	Kappa	**.664**	.083	7.160	.000
N of Valid Cases		60			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

The final step towards validating the Task Analysis rubric was to perform an item

analysis. Scores on both the pre and post-assessment showed normal distribution

patterns (see Appendix G) with two notable exceptions. In the pre-assessment

responses for question 1 (What did you need to know in terms of content?) and

question 5 (What did you need to know in terms of language?) responses were more

heavily weighted towards the low end of the scale, suggesting a "floor effect," or an

inability of the rubric to accurately measure responses at the lower end of the scale.

However, in both questions the post-assessment responses showed normal distribution,

indicating that as participant understanding grew, the rubric was able to capture the

results.

Table 23: Task Analysis Item analysis

| | Paired Differences | | | | | | | Sig. (2-tailed) |
| | Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference | | t | df | |
				Lower	Upper			
Q1	-.62000	1.10454	.15620	-.93391	-.30609	-3.969	49	.000
Q2	-.40000	.98974	.13997	-.68128	-.11872	-2.858	49	.006
Q3	-.22000	.93219	.13183	-.48493	.04493	-1.669	49	.102
Q4	-.30000	.76265	.10785	-.51674	-.08326	-2.782	49	.008
Q5	-.54000	.90824	.12844	-.79812	-.28188	-4.204	49	.000
Q6	-.52000	.76238	.10782	-.73667	-.30333	-4.823	49	.000

A second item analysis was performed to determine if growth was reliably

captured by the rubric. A paired *t* test was run on each question, comparing pre and

post-assessment responses. Each question was shown to measure growth that was

highly statistically significant (see below) with p values below 0.05, with one exception.

Question 3 (What did you need to know in terms of analytical skills?) did not show

statistically significant growth, with a p value of 0.1. Possible explanations for this will be

presented in the Discussion section.

Discussion

      This research was conducted to assist the MSP NGSS K-6 Instructional Specialist grant program in their work around measuring possible shifts in participant understanding of science content and its role in language acquisition through the use of a Task Analysis process during 8-week NGSS summer courses. The aim of this project was to answer two research questions: How well does the task analysis tool help participants understand the language demands of a science task when used as part of an NGSS summer teacher professional development course? How effective is a work sample and scoring rubric protocol for measuring the impact of the Task Analysis process?

      The results of this project indicate statistically significant growth in teacher understanding of the domains of content, analytical skills and language as defined by the ELPA21 standards and NGSS after engaging in the 8-week NGSS courses that utilized the Task Analysis process. Participant interviews uncovered a range of self-reported readiness with regards to using the Task Analysis tool with colleagues, with 6 of the 7 interviewed participants reporting they would consider using it (or already had in the case of one participant). Overall, results indicated that the inclusion of the Task Analysis process was indeed successful in improving participant understanding of the language demands of a science task, although its success could likely be improved in future iterations with some adjustments to both the tasks used for analysis and the assessment itself (see the Recommendations section).

Despite reports from the course instructors that the use of language was not as highly emphasized in these courses as content knowledge, and that language was a domain they were less confident in themselves as instructors, participant scores showed statistically significant gains in this area. This result suggests that even though course instructors were not language specialists, and not confident in leading participants in discussions about language use, the experience was still effective in increasing participants' ability to demonstrate their understanding of the Task Analysis questions on the post-assessment. Elementary teachers, while often insecure about their science conceptual knowledge, may see themselves as much more proficient in language acquisition, and they likely have language expertise to bring to the discussion (Zweip et al., 2011; Banilower et al., 2013). In addition, instructors expressed that they appreciated the opportunity to take the learner stance within the language discussions that happened while using the Task Analysis, much like they ask their participants to take a learner stance while discussing science concepts.

The area that showed the most growth in percent optimal scores was participant understanding of the content and practices knowledge required to perform the task, with a 22% increase in scores in the percent optimal range from the pre to post-assessment. Content knowledge was the primary focus of the summer courses, and the area where course instructors felt the most confident in moving their participants forward in their understanding, which may at least partially explain this change. Participants also benefitted from the previous seven weeks' worth of investigations around the science concepts, which may have contributed to a deepened

understanding of the concepts presented in the course.

The domain that showed the weakest growth was analytical skills, which had a gain of 10% of scores in the percent optimal range, and a small effect size from pre to post-assessment. There are a few possible explanations for this finding. One is that analytical skills may not have been discussed as explicitly or covered as deeply throughout the courses, as compared to content and language knowledge. Unfortunately, this study did not collect data on the delivery of the course content, so there is no way to prove whether or not analytical skills received the same amount of focus as the other two domains. Another possibility is that the pre-assessment scores for analytical skills were much higher than the other two domains, with a mean of 2.24 (as opposed to 1.92 in content and 1.65 in language) so there was less "room to grow" in participant post-assessment responses. A final possibility is that distinguishing between what participants needed to know and do in terms of analytical skills may be much more nuanced than in content or language. Question 4, "What did you need to be able to do in terms of analytical skills," was the only question where statistically significant growth was not observed from the pre to post-assessment (see Table 19). In the rubric, the main difference between "know" and "do" was defined as describing a needed skill versus describing engaging in the process of using that skill (see Appendix A), but this distinction is very subtle and may need to be teased out more in the rubric itself, and in guiding the Task Analysis discussions in the future.

When results were broken out by course, there were differences in how participants scored. The participants in Change Over Time showed the largest growth

overall, and their growth in percent optimal scores was particularly strong in the domains of content and language. There are a few possible explanations for this. The instructor of both Change Over Time courses was also the Course Coordinator, responsible for providing guidance on the incorporation of the Task Analysis tool. She likely spent much more time on learning about the Task Analysis process in preparation to support the other instructors, which may have translated into improved instruction. Another possibility is that the tasks she selected lent themselves particularly well to analysis. Finally, both Change Over Time courses engaged in discussion of the Task Analysis process before the post-assessment, which correlated with growth from pre to post-assessment scores.

One particularly interesting result was the differences observed in growth between the course sessions where the Task Analysis was discussed before the post-assessment was administered, and the course sessions where the Task Analysis was not discussed. In the sessions where the process was discussed, overall percent optimal scores rose 21% from pre to post-assessment, as opposed to 4% from pre to post-assessment for the groups which did not engage in discussion. This finding could suggest that the group discussion is an important component to the Task Analysis process, which would not be a surprise as the process is designed to be engaged in collaboratively.

When examining results for the subpopulations of participants, growth was observed across the demographic groups that were examined. Participants showed growth in their overall understanding of the Task Analysis, whether they taught

86

primary grades, upper elementary, or middle school. This indicates that while

application of knowledge could look quite different at different grade levels, teachers of

any grade level can benefit from engaging in the Task Analysis process and thinking

more deeply about the knowledge and skills associated with content, analytical skills

and language.

However, when looking more closely at differences between teachers of

different grade levels, middle school teachers showed the largest growth in content

knowledge (from 7% of scores in the percent optimal range to 43% of scores) and no

growth in percent optimal scores in analytical skills or language (it is worth noting that

the middle school teachers were the smallest group of the three grade level groups,

with only seven teachers). When analyzed for change in scores from the pre to post-

assessment, middle school teachers showed a similar lack of growth (see Appendix I), so

this finding is not simply a result of low scores being missed by the percent optimal

treatment of the results. This study only included seven middle school teachers, so it's

difficult to draw any conclusions. However, for future iterations of the course, it might

be interesting to consider why middle school teachers, who are content specialists,

might score so much lower on the pre-assessment in content, and why they did not

demonstrate the same kind of growth as their primary and upper elementary grade

teacher colleagues.

One important component of the Task Analysis process is to select a task for

analysis. This proved to be especially challenging in the context of this summer course,

with the knowledge that the task selected for analysis could very much impact the

results. Instructors collaborated on selecting tasks, with some guidance from the course coordinator and myself. This was definitely a learning experience for all of us. One focus was the "grain size" - was the task the right size to analyze? If a task is too limited in scope, it would not provide the opportunity for rich discussion. On the other hand, if a task is too broad, responses to it may be scattered enough that the discussion yields a similar lack of depth. The goal of the Task Analysis process is ultimately to uncover language demands that students may be grappling with, and to determine what supports and scaffolds they would need in order to both access and successfully complete the task. In the Recommendations section, I go into more detail about what we learned through this process around selecting a task for analysis, as well as recommendations for how best to approach choosing a task for future iterations of the course.

This project utilized a tool that was already created, the Task Analysis tool, in a new way: to capture data on any demonstrated changes in understanding of the knowledge and skills needed in content, analytical skills and language. In order to capture this data, a rubric was created and refined, with the goal of measuring any change in teacher understanding of these domains as they are defined by the NGSS and ELP Standards. The rubric addressed the first research question - How well does the ELPA21 Task Analysis tool help participants understand the language demands of a science task when used as part of an NGSS summer professional development course? - by providing quantitative data that could then be analyzed and discussed. The grant

staff was able to look at the findings of this study and make decisions about future planning for the next round of NGSS courses.

The process of creating and validating the rubric and scoring responses addressed the second research question - How effective is a work sample and scoring rubric protocol for measuring the impact of the Task Analysis process? The rubric reliably captured growth in all six assessment questions except for one, as discussed previously. Growth was observed across demographic groups, and across courses, despite differences in participant experiences in language and science prior to the courses, as well as differences in how the materials were presented in each course. The process of creating the rubric also provided more guidance to the course instructors who participated, which will serve to better define the language component in the next iteration of summer courses.

The addition of the Task Analysis process to five 8-week summer NGSS courses highlighted some of the challenges and benefits of shifting how educators approach both science conceptual understanding and second language acquisition. The literature surrounding the implementation of the Next Generation Science Standards and the English Language Proficiency standards agrees that in order to adequately prepare teachers for the challenges of shifting their instruction, sustained professional opportunities are needed (Lee et. al. 2013; Walqui & Heritage, 2012; NRC, 2011; Santos et. al. 2012). In this study, we found participant responses often placed emphasis on content specific vocabulary over conceptual understanding. When asked "In order to successfully complete this task, what did you need to know in terms of content?" 28%

of responses included a mention of vocabulary in the pre-assessment, and 26% in post-assessment responses. This is striking, because while content-specific vocabulary is certainly an important consideration in learning new science content, the ability to use content-specific vocabulary is not necessarily an indicator of conceptual understanding (Brown & Ryoo, 2008). In fact, teacher participants may have been better able to explain a concept in their own words without complex domain-specific vocabulary. When considering English learners in the classroom, this over-emphasis on vocabulary may put all students, and particularly ELs, at a disadvantage as they will face greater challenges in demonstrating their science understanding *if* science understanding is largely equated with content specific vocabulary and language (Brown & Ryoo, 2008). These findings indicate that the importance of conceptual understanding in science learning must continue to be emphasized over the memorization of vocabulary words. It will be recommended that course instructors place more emphasis on accepting "flawed" language as a means to describe conceptual understanding, both from course participants themselves, and from students in the classroom.

Similarly, an emphasis on vocabulary was observed when participants responded to the question "In order to successfully complete this task, what did you need to know in terms of language?" In the pre-assessment responses, the majority of responses mentioned vocabulary, with over half of responses only discussing a grasp of the vocabulary as the language needed to complete the task. Each task involved an interactive language component, such as "discuss the patterns you observe," "develop and explain a model [with a small group]," "orally compare and critique the following 90

arguments," etc. However, very few participants discussed the skills they needed to perform these tasks, such as an understanding of how to use the connecting words needed to discuss patterns, or how to express ideas when there is disagreement in the group around how to represent a phenomenon. This indicates that many teachers may still be struggling with thinking beyond vocabulary when considering what kind of language development occurs in the science classroom, and what kinds of language barriers ELs may be facing. These findings may also point to the prevalence of vocabulary-centered professional development and curriculum that has, until recently, been the main focus of preparing teachers to support language acquisition in the science classroom (CSSO, 2012; Lee, Quinn & Valdés, 2013). Suggestions of how to address this emphasis on vocabulary will be discussed further in the Recommendations section.

Another commonly cited challenge in implementing any new standards initiative is time. The course instructors shared this concern as they planned to incorporate the Task Analysis process into their courses. There simply was not enough time for the instructors to learn about the process, for them to meet and plan how to best use it in their courses, and for the participants to engage with the process during the course. However, even with limited time on all ends, growth was shown, making the argument that the process was worthwhile. There are plans to incorporate the Task Analysis into the next iteration of NGSS courses, with the added benefit that both instructors and participants will be better versed in the process. Perhaps this process has shown that

91

while time can be a limiting factor in professional development, there is value in trying

new approaches and learning as you go.

Limitations

This study had some limitations, which must be considered when looking at the

results. This study observed the results of an intervention into three different courses,

and six different course sessions taught by five instructors. The courses, with the

exception of the administration of the post-assessment, were not observed by the

researcher, meaning that any key differences in how the material was presented were

not included in the data.

Each of the three courses had its own task for analysis, which, coupled with the

lack of data surrounding how the tasks and the Task Analysis process were presented to

participants, affected how well we could compare the results from different courses.

Instructors also reported different levels of readiness to introduce the Task Analysis

process and guide discussions around it, which may have added further uncontrolled

variables to the study. Therefore, we are not really comparing apples and apples when

we compare results from two different courses, because participants were working with

two different tasks which may not have been equally suited to the purpose of analysis,

and their experiences of learning about the Task Analysis process may have been very

different.

Another limitation of this study is that it focused on participant self-reflection,

rather than classroom practice. Participants were asked to think about the skills and

knowledge *they* needed to complete the task, not what their students would need to   92

know or do. We felt that this self-reflection was a critical first step to then start thinking about student outcomes and what kinds of linguistic and cognitive demands they would experience within a science task. However, the literature on science professional development emphasizes the importance of classroom-based practice and ongoing coaching for teachers, neither of which were the focus of this intervention (Penuel et al., 2007, Adamson, Santau & Lee, 2013).

In designing the rubric, I asked instructors to score their own participants' responses as part of the validation process. However, in hindsight, it may have been more beneficial to have instructors score responses from participants who were not part of their course. I found that instructors brought their knowledge of what was covered within the course to their scoring, and as a result, most scoring discrepancies indicated that instructors scored their participants lower than I did. Scoring responses from other courses might have resulted in a more objective process that would be better focused on the rubric design itself, rather than how the responses reflected the course instructor's experience.

Recommendations

The Task Analysis process shows promise as a way to emphasize science and language educational shifts within a science professional development setting. The tool functioned well in a science context, particularly when used to analyze NGSS-aligned tasks, which should incorporate language and analytical skills into a science or engineering task through the Science and Engineering Practices. Beyond the ability of the Task Analysis to highlight the language and cognitive demands of a science task for

teachers, this tool could have the added benefit of "making the case" for how science

instruction can simultaneously achieve instructional goals in language acquisition, math,

English language arts, etc. Previous professional development studies have shown that

when language is highlighted and emphasized within science professional development,

it can lead to more staff buy-in for devoting instructional minutes to science within the

elementary school day and even increased test scores for students (Zweip et al., 2011;

Carrejo & Reinhartz, 2014). Many of the participants interviewed for this project

mentioned that their time for science is very limited, due to factors such as block-

scheduled school days, administrators who don't see the value of prioritizing science

instructional time, and the pressures of standardized testing. The Task Analysis tool

could be used as evidence of how much language development can be present in

science tasks when discussing school day priorities with administrators.

The Task Analysis process is at its most useful when a task is carefully selected

for the purpose of analysis. This process of selecting an appropriate task would certainly

benefit from iteration; as more experience is gained in using the Task Analysis process,

instructors would become more skilled at selecting an appropriate task. Our instructors

grappled with this last summer, as the process was new to all of us and we had not yet

tried it in a professional development setting. My observations after reading through

participant responses were that participants placed a lot of emphasis on the vocabulary

of the task, and not nearly as much on the highlighted practice. I would encourage those

who want to use the Task Analysis in a science professional development setting to

think about highlighting and focusing on the language-rich practice of the task, to

direct the thinking towards the language demands of engaging in the NGSS Practices.

For example, if the task involves modeling a phenomenon or design, the part of the task

that has to do with the modeling practice could be written in bold, and participants

could be asked to focus on the bolded part of the task when responding (see below).

The initial task from the Change Over Time course was the following:

Task 1
Working in a Small Group:
Observe the slinky at rest. What is going on? Use pictures and words to develop a model of
the system.

Move the slinky from side to side across the floor, first slowly and then quickly. What is going
on? Use pictures and words to develop a model that describes the relationship between the
rate at which you move the slinky side to side across the floor and the resulting properties of
the wave you create.

We found that when responding to this task, many participants emphasized content

specific vocabulary in their responses in the domains of content and language (see

Appendix H). In the next iteration of these courses, we wondered about how we could

edit the task to elicit responses that focused more on modeling and a conceptual

understanding of waves.

Modeling is an NGSS Practice which is particularly language-rich (Quinn, Lee &

Valdés, 2012), and therefore can be a fruitful practice to emphasize within a task. As

written, however, this task may be too large of a "grain size," meaning it may involve

too many components. Participants are asked to analyze the process of two different

models when thinking about the task. However, the content knowledge, language and

analytical skills needed to complete each model may have been different, which would

not allow for the amount of depth that we could have seen had the task been more

95

focused. Furthermore, creating a model of waves involved a lot of content-specific

vocabulary, which may explain how often participants referenced "vocabulary" when

responding to this task. Vocabulary could be de-emphasized either in the wording of the

task, or through how the task is introduced to participants.

The inclusion of "Working in a small group" is an important one, because the

collaborative aspect of the task is where much of the language happens. When

participants discuss and collaboratively create a model, they must engage in

argumentation, explanation and negotiation, skills which revolve around an

understanding of the science concept, language use and analytical skills. Framing tasks

in a collaborative way is supported in the literature, which emphasizes the importance

of language learning that happens in social contexts while tackling relevant content-

based tasks (Walqui & Heritage, 2012; Quinn, Lee & Valdés, 2012).

This task from the Change Over Time course could be re-written as such:

Revised Task 1
Working in a Small Group:
Move the slinky from side to side across the floor, first slowly and then quickly. What is going on?
Use pictures and words to develop a model that describes the relationship between the rate at which you move the slinky side to side across the floor and the resulting properties of the wave you create. *Use the language you are most comfortable with to describe what is happening.*

In this way, the context of the task can still be present, but the language-rich practice of

modeling would be more greatly emphasized in how the participant thinks about the

task as a whole. Editing the task to a smaller grain size could help to focus responses

more on the language-rich modeling practice. Bolding the part of the task that is

directly related to modeling emphasizes the response around the language-rich practice.

The addition of the last sentence in italics may de-emphasize the content-specific

vocabulary of waves in order to push participants to think more deeply about the

language-rich aspects of creating the model, such as the discussion that had to happen

in their small group to reach a consensus. This could also provide an opportunity for

future revisions of the model - participants could add more content-specific vocabulary

as they attach it to their conceptual understanding.

The instructor from Change Over Time had similar feelings about how to modify

this task for the next iteration. She expressed that she would cut out the first part of the

original task when using it in the future, not only because the grain size would be more

appropriate for analysis, but also because the second part of the task deals with the

conceptual understanding of forces, which is the content most relevant to the course.

She felt that the collaborative nature of creating a model pushed participants to talk

more about science as a process, rather than just a body of knowledge, which was one

of her instructional goals.

Participant responses may also benefit from an assessment that is more

supportive and directed. For example, instead of asking "In order to successfully

complete this task, what did you need to know in terms of language?" participants could

be asked more specific questions, such as: "What language forms and functions did you

need to know in order to complete the task? How did you use these language forms and

functions to successfully complete the task?" As part of their work with the Task

Analysis, participants would be provided with supports that define language forms and₉₇

functions (see Appendix J). Another example of a modification to the assessment could be to ask participants to analyze the use of register required for the task, such as: the register of scientific argumentation within the classroom, or the register of a popular science article (Lee, Quinn & Valdes, 2013). Emphasizing registers also opens the conversation to the more social qualities of language development in science (Turkan et al., 2014).

The newly developed Task Analysis Rubric could be useful in future professional development settings. Although it is designed to be used with a science task, the rubric could be edited to reflect the content and practices in other content areas, such as English language arts and math. As written, the rubric could be used in the future to measure the effect of using the Task Analysis process in a professional development setting, much as it was used in this study. The rubric could also be used as a guiding document to help teachers better understand what the Task Analysis tool is asking them to do, or as a reflection piece for teachers using the tool as part of a professional learning community. Ideally, the Task Analysis process would be used to facilitate deep professional conversations and planning, not to rate performance, so using the rubric to score teacher responses in the context of the school building is not recommended.

The rubric as written briefly touches on the Crosscutting Concepts, which can bring to light some interesting discussions. Depending on how Crosscutting Concepts are emphasized within a task, they could fall into the Analytical Skills category (ex: examining seasonal change through a Cause and Effect lens) or content (ex: seasonal change is caused by the tilt of the Earth's axis). Discussions like these could help

teachers deepen their understanding of both the domains of the Task Analysis and the Crosscutting Concepts themselves, but they may not naturally occur unless Crosscutting Concepts are more explicitly called out in the task for analysis, or in the tools themselves.

The Task Analysis rubric, like any rubric, is a work in progress. It served the purposes of the NGSS summer course evaluation, but would need to be edited for use in different contexts. Ideally, such editing, much like the Task Analysis process itself, would be happen in a collaborative setting. In collaboratively editing this or any rubric, professional developers make time to come to common understandings of what exactly these new standards are expecting of teachers and their students, and how they can best meet the needs that arise.

Finally, although the Task Analysis rubric was created with the hope that it could be used again in the future, there is the very real challenge of the amount of time that is required to score participant responses. Without the help of a graduate student, it would be exceedingly difficult to score future responses (it is certainly outside of the scope of the course instructors to do so). Therefore, it may be necessary to randomly select a subset of responses to score, which, while not providing the same volume of data, would provide some insight into how participant understanding of the Task Analysis process may be progressing from one iteration of the courses to the next. Other professional development providers may encounter the same challenges of time, and may choose instead to use the rubric in some of the other ways that are recommended,

such as a self-reflection piece, or a guiding document to be used with the Task Analysis

tool.

References

ACT. (2012). *The Condition of College & Career Readiness.* Retrieved on October 18,

    2016 from http://media.act.org/documents/CCCR12-NationalReadinessRpt.pdf/

Adamson, K., Santau, A., & Lee, O. (2013). The Impact of Professional Development on

    Elementary Teachers' Strategies for Teaching Science with Diverse Student

    Groups in Urban Elementary Schools. *Journal Of Science Teacher Education*,

    24(3), 553-571.

Anstrom, K., DiCerbo, P., Butler, F., Katz, A., Millet, J., & Rivera, C. (2010). *A review of*

    the literature on Academic English: Implications for K–12 English Language

    Learners. Arlington, VA: The George Washington University Center for Equity and

    Excellence in Education.

Bailey, A. L., & Butler, F. A. (2003). *An evidentiary framework for operationalizing*

    academic language for broad application to K–12 education: A design document.

    Los Angeles: University of California, National Center for Research on Evaluation,

    Standards, and Student Testing (CRESST).

Banilower, E. R., Smith, P. S., Weiss, I. R., Malzahn, K. A., Campbell, K. M., & Weis, A. M.

    (2013). *Report of the 2012 national survey of science and mathematics*

    education. Chapel Hill, NC: Horizon Research, Inc.

Block, D. (2003). The social turn in second language acquisition. Washington, D.C.:

    Georgetown University Press.

Brown, B., & Ryoo, K. (2008). Teaching Science as a Language: A 'content-first'

approach to science teaching. *Journal of Research in Science Teaching*, 45(5), 529-553.

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. *Colorado Springs, CO: BSCS*, *5*, 88-98.

Carrejo, D. J., & Reinhartz, J. (2014). Teachers fostering the co-development of science literacy and language literacy with English language learners. *Teacher Development*, *18*(3), 334-348.

Cheuk, T. (2013). *Relationships and convergences among the mathematics, science, and ELA practices.* Refined version of diagram created by the Understanding Language Initiative for ELP Standards. Palo Alto, CA: Stanford University.

Council of Chief State School Officers. (2012). *Framework for English Language Proficiency Development Standards Corresponding to the Common Core State Standards and the Next Generation Science Standards*. Retrieved from http://www.ccsso.org/Documents/2012/ELPD%20Framework%20Booklet-Final%20for%20web.pdf

Council of Chief State School Officers. (2014). *English Language Proficiency (ELP) Standards*. Retrieved from http://www.ode.state.or.us/opportunities/grants/nclb/title_iii/final-4_30-elpa21-standards.pdf

CTB/McGraw-Hill for the Oregon Department of Education. (2007). *Oregon English Language Proficiency Assessment (ELPA) Proficiency Level Descriptors (PLDs).*

Retrieved October 17, 2016 from https://www.4j.lane.edu/wp-content/uploads/2013/03/4j_instruction_ELPA_proficiency.pdf

ELP Standards: Overview. (n.d.). Retrieved October 19, 2016, from

http://www.elpa21.org/sites/default/files/ELP_Standards_Overview.pdf

Firth, A., & Wagner, J. (1997). On discourse, communication, and (some) fundamental

concepts in SLA research. The Modern Language Journal, 81(3), 285-300.

Fleischman, H., Hopstock, P., Pelczar, M. & Shelley, B. (2010). *Highlights from PISA*

2009: Performance of U.S. 15-year-old students in reading, mathematics, and

science literacy in an international context. National Center for Educational

Statistics. Retrieved October 18, 2016 from:

http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2011004

Grossman, P. L. (1990). The making of a teacher: Teacher knowledge and teacher

education. New York: Teachers College Press.

Gumperz, J. & Levinson, S. (Eds.). (1996). Rethinking linguistic relativity. Cambridge, UK:

Cambridge University Press.

Hattie, J. & Timperley, H. (2007). The power of feedback. Review of Educational

Research, 77(1), 81-112.

Lakshmanan, A., Heath, B. P., Perlmutter, A. & Elder, M. (2011). The Impact of Science

Content and Professional Learning Communities on Science Teaching Efficacy

and Standards-Based Instruction. Journal Of Research In Science Teaching, 48(5),

534-551.

Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for

categorical data. *Biometrics* 33:159-174.

Lantolf, J. & Thorne, S. (2006). Sociocultural theory and the genesis of second language

development. Oxford: Oxford University Press.

Lee, O., Llosa, L., Jiang, F., Haas, A., O'Connor, C. & Van Booven, C. (in press). Teachers'

science knowledge and practices with English language learners. *Journal of

Research in Science Teaching.*

Lee, O., Miller, E. C. & Januszyk, R. (2014). Next Generation Science Standards: All

Standards, All Students. *Journal Of Science Teacher Education*, 25(2), 223-233.

Lee, O., Quinn, H. & Valdés, G. (2013). Science and language for English language

learners in relation to Next Generation Science Standards and with implications

for Common Core State Standards for English language arts and mathematics.

Educational Researcher, 42(4), 223–233.

Levinson, S. (2003). Space in language and cognition. Cambridge, UK: Cambridge

University Press.

Lindahl, K., & Watkins, N. M. (2015). Creating a Culture of Language Awareness in

Content-Based Contexts. *TESOL Journal, 6*(4), 777-789. doi:10.1002/tesj.223

Lucy, J. (1992). Language diversity and thought: A reformulation of the linguistic

relativity hypothesis. Cambridge, UK: Cambridge University Press.

McMurrer, J. (2008). Instructional time in elementary schools: A closer look at changes

for specific subjects. Washington, DC: Center on Education Policy.

Module Two: Task Analysis Process. (n.d.) Retrieved from

http://prodev.elpa21.org/module2/index.html#01_GuidingQuestionsObjectives

Moll, L., Amanti, C., Neff, D. & Gonzalez, N. (1992). Funds of knowledge for teaching:

Using a qualitative approach to connect homes and classrooms. Theory Into

Practice, XXXI, 2, 132-141.

National Research Council. (2011). *A framework for K-12 science education: Practices,*

crosscutting themes, and core ideas. Washington, DC: National Academies Press.

NGSS Lead States. (2013). *Next Generation Science Standards: For States, By State*s.

Washington, DC: The National Academies Press.

Olsen, L. (2010). Reparable harm: Fulfilling the unkept promise of educational

opportunities for Long Term English Language Learners. Long Beach, CA:

Californians Together.

Oregon Department of Education. (2009). *Oregon Science K-HS Content Standards.*

Retrieved October 18, 2016 from

http://www.ode.state.or.us/teachlearn/subjects/science/curriculum/2009feb-

adopted-k-h-science-standards.pdf

Oregon Department of Education. (2013). *Oregon English Learners Statewide Strategic*

Plan 2013-2016. Retrieved March 17, 2016 from

http://www.ode.state.or.us/superintendent/priorities/final-equity-lens-draft-

adopted.pdf

Oregon Department of Education. (2016) Reports and Data: Information on

Assessments, Curriculums, Funding, Report Cards, and other reports. Retrieved

April 25, 2017 from http://www.oregon.gov/ode/educator-resources/assessment/Pages/Assessment-Group-Reports-for-2014-2015-and-2015-2016.aspx

Park, S., Chen, Y-C. & Jang, J. (2008). *Developing measures of teachers' pedagogical content knowledge for teaching high school biology.* International conference of the Association for Science Teacher Education, St. Louis, MI.

Park, S., Jang, J., Chen, Y. & Jung, J. (2011). Is pedagogical content knowledge (PCK) necessary for reformed science teaching? Evidence from an empirical study. *Research in Science Education*, 41, 245–260.

Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What Makes Professional Development Effective? Strategies That Foster Curriculum Implementation. *American Educational Research Journal, 44*(4), 921-958. doi:10.3102/0002831207308221

Penuel, W. R., Harris, C. J., & DeBarger, A. H. (2015). Implementing the Next Generation Science Standards. *Phi Delta Kappan*, *96*(6), 45-49. doi:10.1177/0031721715575299

Professional Development | Elpa21. (2014). Retrieved June 09, 2016, from http://www.elpa21.org/standards-initiatives/professional-development

Quinn, H., Lee, O. & Valdés, G. (2012). Language demands and opportunities in relation to Next Generation Science Standards for English language learners: What teachers need to know. Stanford, CA: Stanford University, Understanding Language Initiative at Stanford University (ell.stanford.edu).

Rampey, B., Finnegan, R., Goodman, M., Mohadjer, L., Krenzke, T., Hogan, J., Provasnik,

S. & Xie, H. (2016). Skills of U.S. Unemployed, Young, and Older Adults in Sharper

Focus: Results From the Program for the International Assessment of Adult

Competencies (PIAAC) 2012/2014. U.S. Department of Education. Washington,

DC: National Center for Education Statistics. Retrieved October 18, 2016 from

http://nces.ed.gov/pubs2016/2016039rev.pdf/

Riggs, I. & Enochs, L. (1990). Toward the development of an elementary teacher's

science teaching efficacy belief instrument. Science Education, 74, 625–638.

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

Instructional Science, 18, 119- 144.

Sandholtz, J. H. & Ringstaff, C. (2014). Inspiring Instructional Change in Elementary

School Science: The Relationship between Enhanced Self-Efficacy and Teacher

Practices. *Journal Of Science Teacher Education*, *25*(6), 729-751.

Santos, M., Darling-Hammond, L. & Cheuk, T. (2012). *Teacher development to support*

English language learners in the context of common core state standards.

Stanford, CA: Stanford University, Understanding Language Initiative at Stanford

University (ell.stanford.edu).

Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., Bloom, I.

(2002). Measuring reform practices in science and mathematics classrooms: the

reformed teaching observation protocol. *School Science and Mathematics*,

102(6), 245–253.

Shulman, L. (1987). Knowledge and teaching: foundations of the new reform. *Harvard Educational Review*, 57(1), 1–22.

Sullivan, G. M.,= & Feinn, R. (2012). Using Effect Size—or Why the *P* Value Is Not Enough. *Journal of Graduate Medical Education*, *4*(3), 279–282. http://doi.org/10.4300/JGME-D-12-00156.1

The Need for Standards. (n.d.). Retrieved May 12, 2017, from https://www.nextgenscience.org/need-standards

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

Turkan, S., De Oliveira, L. C., Lee, O. & Phelps, G. (2014). Proposing a Knowledge Base for Teaching Academic Content to English Language Learners: Disciplinary Linguistic Knowledge. *Teachers College Record, 116*(030308).

Understanding Language Initiative. (2012). The purpose of English language proficiency standards, assessments, and instruction in an age of new standards: Policy statement from the Understanding Language Initiative. Palo Alto, CA: Author. Retrieved July 29, 2013, from http://ell.stanford.edu/sites/default/files/ELP_task_force_report_rev.pdf

Van Lier, L. (2004). The ecology and semiotics of language learning: A sociocultural perspective. Dordrecht, NL: Kluwer Academic Press.

Walqui, A. & Heritage, M. (2012). *Instruction for Diverse Groups of English Language Learners.* Stanford, CA: Stanford University, Understanding Language Initiative at Stanford University (ell.stanford.edu).

Zwiep, S. G., Straits, W. J., Stone, K. R., Beltran, D. D., & Furtado, L. (2011). The

Integration of English Language Development and Science Instruction in

Elementary Classrooms. *Journal of Science Teacher Education, 22*(8), 769-785.

doi:10.1007/s10972-011-9254-z

Appendices

Appendix A: Task Analysis Rubric

<table>
<tr><td colspan="5" align="center">Task Analysis Rubric
for use with science-based tasks</td></tr>
<tr><td></td><td>1 - limited understanding</td><td>2 - emerging understanding</td><td>3 - proficient understanding</td><td>4 - exemplary understanding</td></tr>
<tr><td rowspan="4">Content knowledge</td><td colspan="4" align="center">What did you need to know to successfully complete this task?</td></tr>
<tr><td>Lists content specific vocabulary, skills, and/or scientific facts that may or may not directly relate to the task. No explanation or examples are given.
Answer is too brief to analyze.</td><td>Lists and/or briefly describes science concepts and or skills that relate to the general science of the task
If content specific vocabulary is listed, connections to a science concept(s) are made</td><td>Demonstrates understanding of science concepts by directly relating them to the specific task
Connects content to Practices, Crosscutting Concepts, and/or sense making</td><td>Demonstrates detailed understanding of science concepts that directly relate to the specific task by providing example(s).
Strong connections to practices, Crosscutting concepts, and/or sense making are made</td></tr>
<tr><td colspan="4" align="center">What did you need to be able to do to successfully complete this task?</td></tr>
<tr><td>No practices mentioned, or practices listed do not relate to the task
Answer is too brief to analyze</td><td>Practices listed relate to the task in a generalizable way.
Limited understanding of how science content knowledge is connected to the SEPs associated with the task</td><td>Discusses how science content knowledge is used to perform the practices that are specific to the task
Connections between Practices, Crosscutting Concepts and/or sense making</td><td>Demonstrated detailed understanding of how science content knowledge is used to perform the practices that are specific to the task.
Provides concrete examples that show connection between task and practice.
Strong connections between Practices, Crosscutting Concepts and/or sense making.</td></tr>
</table>

<table>
<tr><td rowspan="4">Analytical skills</td><td colspan="4" align="center">What did you need to know to successfully complete this task?</td></tr>
<tr><td>No analytical skills listed, or skills listed are not appropriate for the task
Answer is too brief to analyze</td><td>Analytical skill(s) listed are appropriate for the task in a generalizable way
Analytical skill(s) provided may not reflect the complexity of the task
Analytical skill(s) demonstrate little connection to content/ practices, Crosscutting Concepts and/or language</td><td>Analytical skill(s) described are appropriate for the task
Analytical skill(s) provided reflect the complexity of the task.(ie. more complex tasks require more complex description of analytical skills needed)
Some connections to content/ practices, Crosscutting Concepts and/or language</td><td>Analytical skill(s) are described in detail and are explicitly appropriate for the specific task
Connections to content/ practices, Crosscutting Concepts and/or language are described in detail.</td></tr>
<tr><td colspan="4" align="center">What did you need to be able to do to successfully complete this task?</td></tr>
<tr><td>No analytical process described, or process is not appropriate for the task
Answer is too brief to analyze</td><td>Analytical process(es) listed are appropriate for the task in a generalizable way
Analytical process(es) provided may not reflect the complexity of the task
Little connection to content/practices, Crosscutting Concepts and/or language is made and described</td><td>Analytical process(es) provided reflect the complexity of the task (i.e. more complex tasks require more complex description of analytical skills needed)
Some connections to science content/practices, Crosscutting Concepts and/or language are made and described</td><td>Analytical process(es) are described in detail and are explicitly appropriate for the specific task
Connections to science content/practices, Crosscutting Concepts, and/or language are made and described in detail</td></tr>
</table>

Language	What did you need to know to successfully complete this task?			
	• Vague reference to content specific vocabulary that was used within the task • No language forms or functions mentioned • No connections made between language and content knowledge • Answer is too brief to analyze	• Language forms and functions are listed in a generalizable way • Limited connections made between language, practice(s), content knowledge and/or and analytical skill(s) (when appropriate) • Limited connection made between content specific vocabulary and conceptual understanding	• Language forms and functions are listed with connection to the task • Connections made between language forms and functions, practice(s), content knowledge and/or analytical skill(s) (when appropriate)	• Language forms and functions are listed with examples of how they were connected to the task • Language forms within the task are discussed in service of language functions • Strong connections made between language and content knowledge
	What did you need to be able to do to successfully complete this task?			
	• No mention of modalities, or language does not relate to the task • Answer is too brief to analyze	• Productive, receptive and/or interactive uses of language are listed in a generalizable way • Limited connections made between language modalities, practice(s), content knowledge and/or and analytical skill(s) (when appropriate)	• Productive, receptive and/or interactive uses of language are listed with connection to the task • Connections made between language modalities, practice(s), content knowledge and/or analytical skill(s) (when appropriate)	• Demonstrated understanding of productive, receptive and interactive uses of language within the task through specific examples • Strong connections made between language modalities, practice(s), content knowledge and/or analytical skill(s) (when appropriate)

For more information on the Task Analysis process, please visit: ELPA 21 Professional Development Module Two: Task Analysis Process

Appendix B: Task Analysis Tool

Task Analysis Tool
Understanding Language Initiative, Stanford University
Please "make a copy" before editing

Name of Task: **Grade Level:**   **Subject:**

Task Analysis Step	Guiding Questions and Resources	Analysis
Step 1: Examine and Identify Appropriate Instructional Task	**Guiding Questions:** Is/does this task: □ Clear in its expectations? □ Grade-level appropriate? □ Aligned to the standards? □ Require students to use language *and* analytical skills to demonstrate their content knowledge?	Is this an appropriate task for analysis? Why?
Step 2: Identify Task Demands	**Guiding Questions:** • Write down everything that students need to demonstrate, know, or do in order to successfully complete this task. • To do this, read (or watch) the task instructions. **Resources:** • For **Content Knowledge**: Common Core State Standards, Next Generation Science Standards, or other relevant standards (e.g., district, state, etc.) • For **Analytical Skills**: Depth of Knowledge (DOK) Levels • For **Language**: Language Functions and Forms PDF	What do students need to do and know in terms of...? **Content Knowledge** / **Analytical Skills** / **Language**

Task Analysis Step	Guiding Questions and Resources	Analysis
Step 3: Identify Disciplinary Practice(s)	**Guiding Question:** What disciplinary practice(s) are most relevant to this task? **Resources:** • Core Disciplinary Practices PDF • Interactive Correspondence between Practices, Tasks, and Functions PDF	What are the relevant disciplinary analytical practices for this task?
Step 4: Identify ELP Standard(s)	**Guiding Questions:** What English Language Proficiency Standards are reflected in this task? Which of these ELP standards do you feel comfortable measuring or intend to assess? **Resources:** • The ELP Standards • Alternative Organization of Standards • The K-12 Practices Matrix	What are the relevant ELP standards for this task? Please include your reasoning behind selecting these standards. **ELP Standard** / **Your Reasoning**

Appendix C: Pre-assessments

Pre-Assessment: Change Over Time
Please complete the following information:
1. Name of your District:
2. Name of your School:
3. Grade level (check all that apply):
    a. K
    b. 1
    c. 2
    d. 3
    e. 4
    f. 5
    g. 6
    h. Other (please explain): _____
4. What is your current role at your school (check all that apply)?
    a. Classroom teacher
    b. ESL teacher
    c. Content specialist/Teacher on Special Assignment
    d. Other (please explain):
5. Do you work with ELLs at your site?
    a. Yes
    b. No
6. If yes, what structures are in place to teach English Language Development with your ELL students? (choose all that apply):
    a. Pull out
    b. Push in
    c. Whole class
    d. Other (please explain):
7. Not including this course, how many science professional development sessions have you participated in within the last 12 months?
    a. 0
    b. 1-2
    c. 3-4
    d. 5+
8. Of these sessions, how many focused on the NGSS?
    a. 0
    b. 1-2
    c. 3-4
    d. 5+
9. Not including this course, how many English language learner-focused professional development sessions have you participated in within the last 12 months?
    a. 0
    b. 1-2
    c. 3-4

d. 5+
10. Of these sessions, how many focused on the ELP standards?
        a. 0
        b. 1-2
        c. 3-4
        d. 5+
11. How confident do you feel about teaching NGSS-aligned science to English language learners?
        1 Not at all confident
        2 Not very confident
        3 Somewhat confident
        4 Fairly confident
        5 Very confident

We have selected a NGSS-aligned task that you engaged with today. Please read it carefully and answer the following questions. As you answer, consider what you needed to know and be able to do to successfully complete this task. It is okay to include the same information in more than one of your responses.

Working in a Small Group:
- Observe the slinky at rest. What is going on? Use pictures and words to develop a model of the system.
- Move the slinky from side to side across the floor, first slowly and then quickly. What is going on? Use pictures and words to develop a model that describes the relationship between the rate at which you move the slinky side to side across the floor and the resulting properties of the wave you create.

In order to successfully complete this task...

1. What did you need to know in terms of science knowledge?
2. What did you need to be able to do in terms of science knowledge?

3. What did you need to know in terms of analytical skills (problem solving)?
4. What did you need to be able to do in terms of analytical skills (problem solving)?

5. What did you need to know in terms of language?
6. What did you need to be able to do in terms of language?

Pre-Assessment: Energy and Matter

Please complete the following information:

1. Name of your District:

2. Name of your School:

3. Grade level (check all that apply):
- a. K
- b. 1
- c. 2
- d. 3
- e. 4
- f. 5
- g. 6
- h. Other (please explain): _____

4. What is your current role at your school (check all that apply)?
- a. Classroom teacher
- b. ESL teacher
- c. Content specialist/Teacher on Special Assignment
- d. Other (please explain):

5. Do you work with ELLs at your site?
- a. Yes
- b. No

6. If yes, what structures are in place to teach English Language Development with your ELL students? (choose all that apply):
- a. Pull out
- b. Push in
- c. Whole class
- d. Other (please explain):

7. Not including this course, how many science professional development sessions have you participated in within the last 12 months?
- a. 0
- b. 1-2
- c. 3-4
- d. 5+

8. Of these sessions, how many focused on the NGSS?
- a. 0
- b. 1-2
- c. 3-4
- d. 5+

9. Not including this course, how many English language learner-focused professional development sessions have you participated in within the last 12 months?
- a. 0

b. 1-2
c. 3-4
d. 5+

10. Of these sessions, how many focused on the ELP standards? _____
     a. 0
     b. 1-2
     c. 3-4
     d. 5+

11. How confident do you feel about teaching NGSS-aligned science to English language learners?
     1 Not at all confident
     2 Not very confident
     3 Somewhat confident
     4 Fairly confident
     5 Very confident

We have selected a NGSS-aligned task that you engaged with today. Please read it carefully and answer the following questions. As you answer, consider what you needed to know and be able to do to successfully complete this task. It is okay to include the same information in more than one of your responses.

Your Task:

Work collaboratively in a small group to explore how adding mass to your cart affects its rate of movement down the ramp. Discuss the patterns you observe and what they tell you about the relationship between mass and the resulting movement of the cart. Use your group data to explain the relationship between mass and speed as it is seen in this system.

In order to successfully complete this task...

1. What did you need to know in terms of science knowledge?
2. What did you need to be able to do in terms of science knowledge?

3. What did you need to know in terms of analytical skills (problem solving)?
4. What did you need to be able to do in terms of analytical skills (problem solving)?

5. What did you need to know in terms of language?
6. What did you need to be able to do in terms of language?

Pre-Assessment: Systems and Interactions

Please complete the following information:

1. Name of your District:

2. Name of your School:

3. Grade level (check all that apply):
    a. K
    b. 1
    c. 2
    d. 3
    e. 4
    f. 5
    g. 6
    h. Other (please explain): _____

4. What is your current role at your school (check all that apply)?
    a. Classroom teacher
    b. ESL teacher
    c. Content specialist/Teacher on Special Assignment
    d. Other (please explain):

5. Do you work with ELLs at your site?
    a. Yes
    b. No

6. If yes, what structures are in place to teach English Language Development with your ELL students? (choose all that apply):
    a. Pull out
    b. Push in
    c. Whole class
    d. Other (please explain):

7. Not including this course, how many science professional development sessions have you participated in within the last 12 months?
    a. 0
    b. 1-2
    c. 3-4
    d. 5+

8. Of these sessions, how many focused on the NGSS?
    a. 0
    b. 1-2
    c. 3-4
    d. 5+

9. Not including this course, how many English language learner-focused professional development sessions have you participated in within the last 12 months?
    a. 0

       b. 1-2

       c. 3-4

       d. 5+

10. Of these sessions, how many focused on the ELP standards? _____

       a. 0

       b. 1-2

       c. 3-4

       d. 5+

11. How confident do you feel about teaching NGSS-aligned science to English language learners?

       1 Not at all confident

       2 Not very confident

       3 Somewhat confident

       4 Fairly confident

       5 Very confident

We have selected a NGSS-aligned task that you engaged with today. Please read it carefully and answer the following questions. As you answer, consider what you needed to know and be able to do to successfully complete this task. It is okay to include the same information in more than one of your responses.

Using pictures and words, develop representations that explain this natural phenomenon:

What is the difference in the relationship between the length in shadows in summer and winter?

- Use the stick and the flashlight to demonstrate what is happening. Represent the position of the sun in winter and summer as you explain the phenomenon.
- Compare the stick and flashlight to show the length of the shadows in winter and summer. What is going on? Use pictures and words to develop and explain a model that describes the relationship between the earth and sun and the length of shadows in summer and winter.

In order to successfully complete this task...

1. What did you need to know in terms of science knowledge?
2. What did you need to be able to do in terms of science knowledge?

3. What did you need to know in terms of analytical skills (problem solving)?
4. What did you need to be able to do in terms of analytical skills (problem solving)?

5. What did you need to know in terms of language?
6. What did you need to be able to do in terms of language?

Appendix D: Post-Assessment

Post-Assessment: Change Over Time

We have selected a NGSS-aligned task that you engaged with today. Please read it carefully and answer the following questions. As you answer, consider what you needed to know and be able to do to successfully complete this task. It is okay to include the same information in more than one of your responses.

Your Task: In your small group orally compare and critique the following arguments:

A: "Living fossil" is a scientifically accurate term. There are many species alive today that we find in the fossil record and many living fossils share similar characteristics (i.e. low diversity of current species, sometimes only one species left, ancient looking).

B: "Living fossil" is not a scientifically accurate term. The fossil record is incomplete and not all physiological changes that happen over time could be captured during the fossilization process.

Refer to evidence from the slides and resources from our work together to analyze each argument.

<u>In order to successfully complete this task...</u>

1. What did you need to know in terms of science knowledge?
2. What did you need to be able to do in terms of science knowledge?

3. What did you need to know in terms of analytical skills (problem solving)?
4. What did you need to be able to do in terms of analytical skills (problem solving)?

5. What did you need to know in terms of language?
6. What did you need to be able to do in terms of language?

Post assessment: Energy and Matter

We have selected a NGSS-aligned task that you engaged with today. Please read it carefully and answer the following questions. As you answer, consider what you needed to know and be able to do to successfully complete this task. It is okay to include the same information in more than one of your responses.

With a partner:

- Compare and discuss the results of your investigation. With your partner, reflect on what you have learned about photosynthesis and the flow of energy and cycling of matter to develop an explanation of your results.

Whole class:

- Using your own personal experiences, data from your investigation, and real life examples, work together as a group to construct an oral explanation of how matter is cycled and energy flows through the process of photosynthesis.

In order to successfully complete this task…

1. What did you need to know in terms of science knowledge?
2. What did you need to be able to do in terms of science knowledge?

3. What did you need to know in terms of analytical skills (problem solving)?
4. What did you need to be able to do in terms of analytical skills (problem solving)?

5. What did you need to know in terms of language?
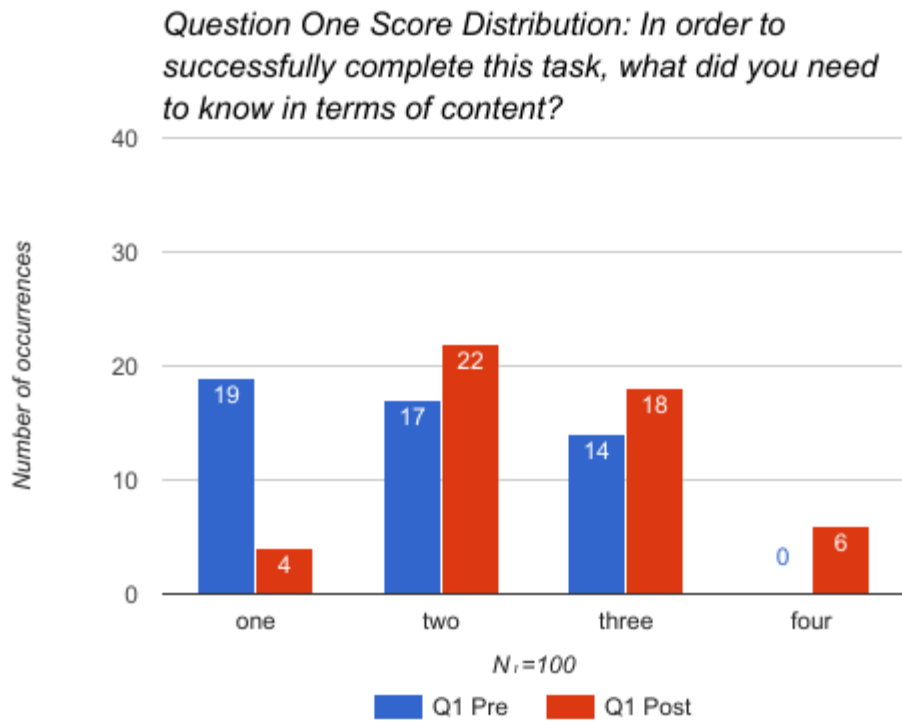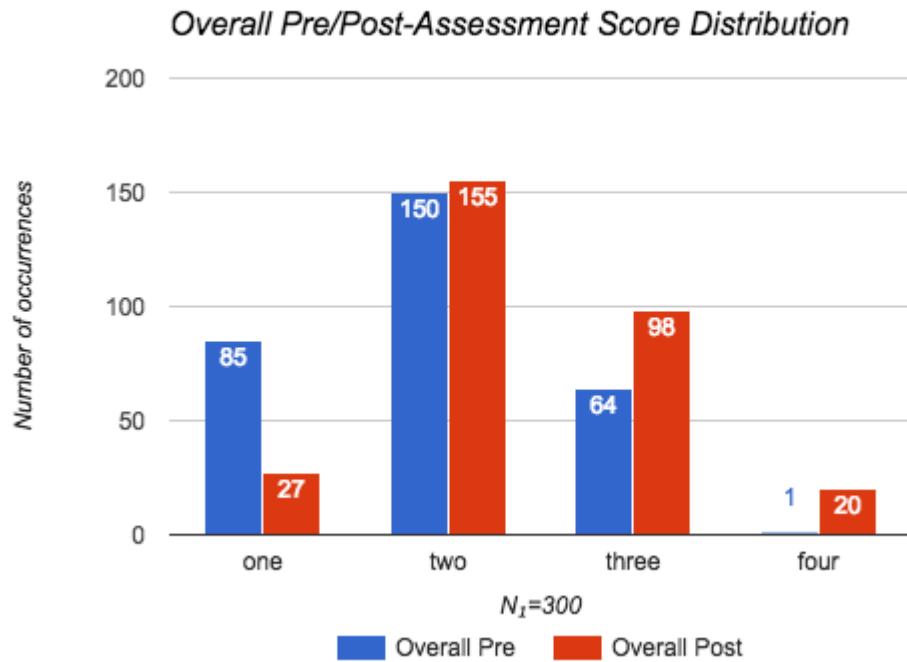6. What did you need to be able to do in terms of language?

Post assessment: Systems and Interactions

We have selected a NGSS-aligned task that you engaged with today. Please read it carefully and answer the following questions. As you answer, consider what you needed to know and be able to do to successfully complete this task. It is okay to include the same information in more than one of your responses.

With your small group: Using your own experience, class data, examples, pictures, and words, develop representations that explain the natural phenomenon associated with the flow of carbon through ecosystems.

- Use a reinforcing causal loop to show the accumulation of carbon dioxide in the atmosphere. A connection circle can be used to trace the flow of carbon dioxide through the ecosystem and create a reinforcing loop, where the amount of carbon dioxide is going up, and up, and up.
- Use a balancing causal loop diagram(s) to show how carbon flows between earth's systems.

In order to successfully complete this task...

1. What did you need to know in terms of science knowledge?
2. What did you need to be able to do in terms of science knowledge?

3. What did you need to know in terms of analytical skills (problem solving)?
4. What did you need to be able to do in terms of analytical skills (problem solving)?

5. What did you need to know in terms of language?
6. What did you need to be able to do in terms of language?

Appendix E: Course participant interview questions

1. What demands for language use do you see the Next Generation Science Standards (NGSS) requiring of students?
2. How well did you understand the demands for language use that the NGSS require of students before taking this course? How well do you understand them now?
3. How prepared do you feel to help students meet language use demands during a NGSS-aligned science lesson?
4. What experiences have you had, in this course or elsewhere, which have prepared you?
5. What kind of support do you still need?
6. What challenges of any kind can you identify that might make it difficult to integrate the English Language Proficiency (ELP) standards into science instruction?
   a. What classroom, school, district, and state level challenges do you anticipate?
7. What supports (district, MSP grant, etc) do you anticipate receiving from your school and/or district to help you implement the NGSS and ELP standards?
8. What school structures are in place (PLCs, collaboration with ESL teachers, etc) that can support you in this work?
9. The design of this course is an iterative process, and feedback will be used to improve future courses. In what ways did the NGSS course meet your needs in relation to teaching NGSS-aligned science to English language learners?
10. In what ways could the course be improved to better meet your needs?
11. What else do you want the MSP grant leadership to know?

Appendix F: Course instructor focus group questions

1. Please share how you structured the task analysis process and reflection in your class.
2. Did the task analysis process go as you had planned?
3. What specifically went well or didn't go well?
4. What instructional decisions did you make in response to what happened during the task analysis process?
5. Did the ELP task analysis process come up again over the 8 weeks?
6. Were the instructor meetings helpful in preparing you to lead the task analysis process? What specifically was helpful? What could have been more helpful?
7. What changes would you make to your course to better support teachers in understanding the language demands of the NGSS?

Overall Pre/Post-Assessment Score Distribution

$N_1=300$

Overall Pre — Overall Post

Question One Score Distribution: In order to successfully complete this task, what did you need to know in terms of content?
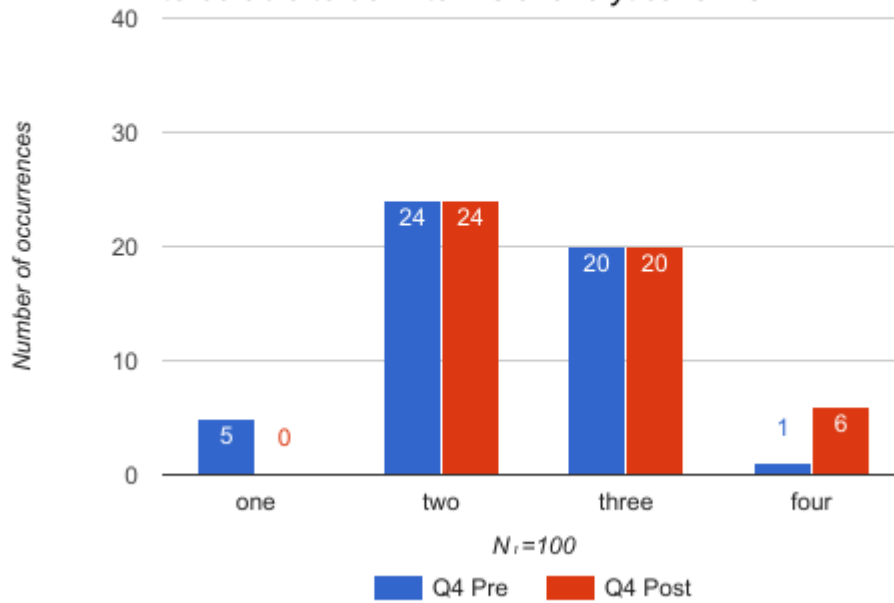

$N_1=100$

Q1 Pre — Q1 Post

Question Two Score Distribution: In order to successfully complete this task, what did you need to be able to do in terms of content?
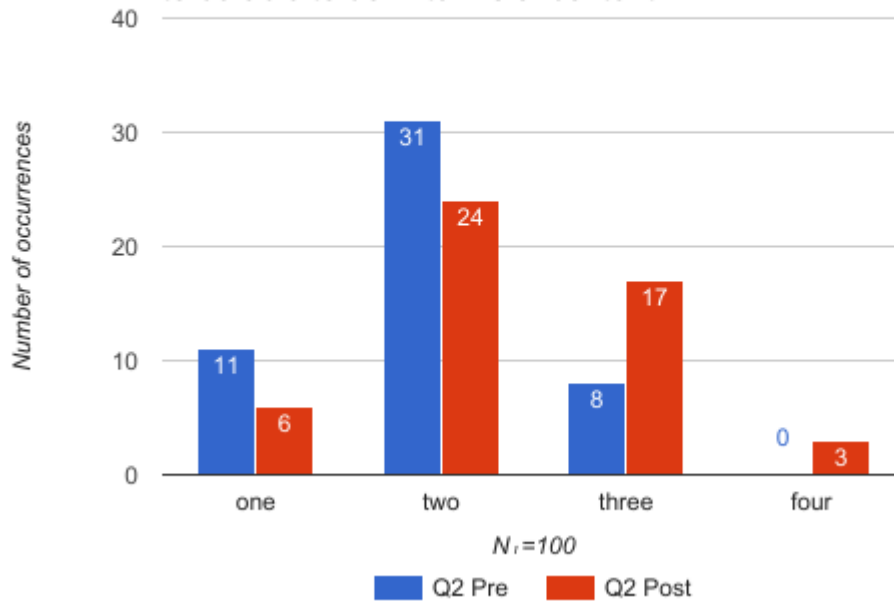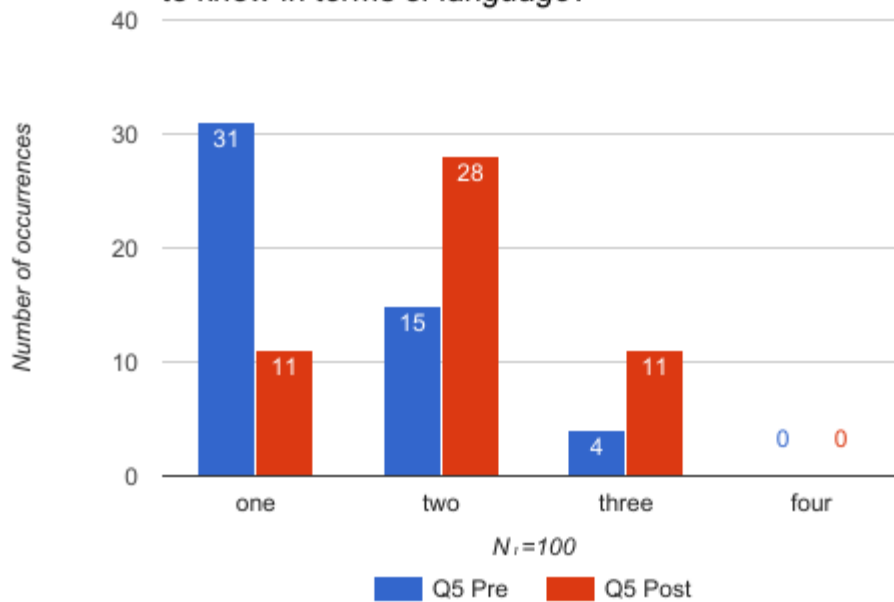

Question Three score distribution: In order to successfully complete this task, what did you need to know in terms of analytical skills?
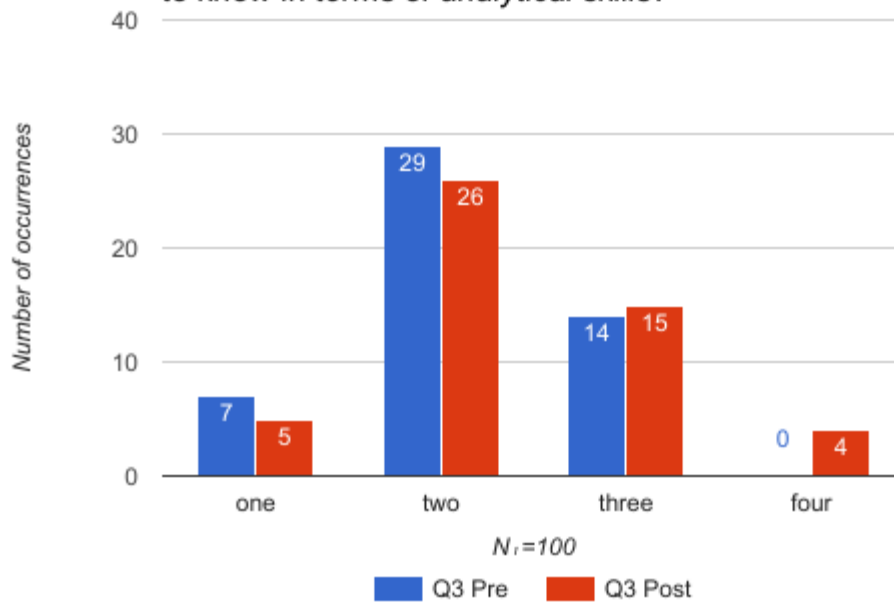
Question Four score distribution: In order to successfully complete this task, what did you need to be able to do in terms of analytical skills?

Question Five score distribution: In order to successfully complete this task, what did you need to know in terms of language?

127

Question Six score distribution: In order to successfully complete this task, what did you need to be able to do in terms of language?
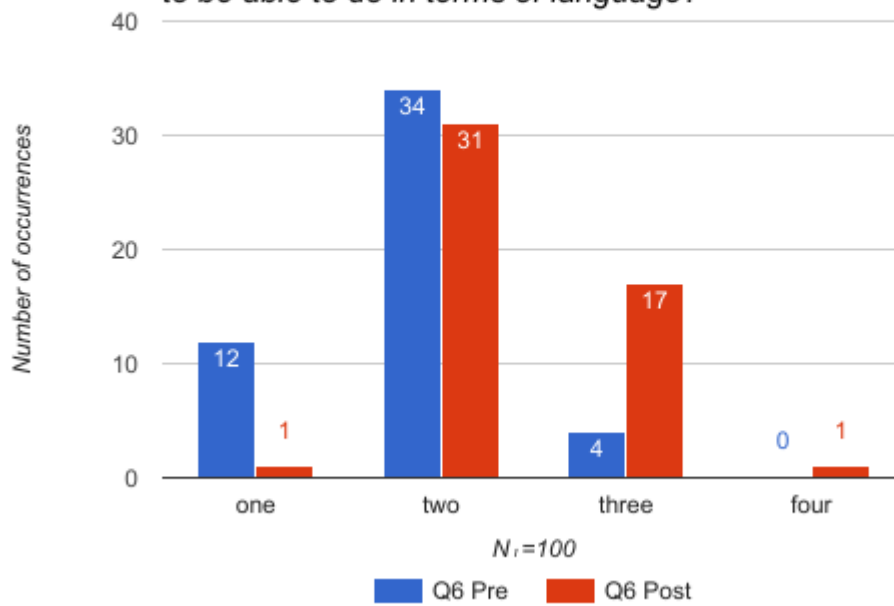
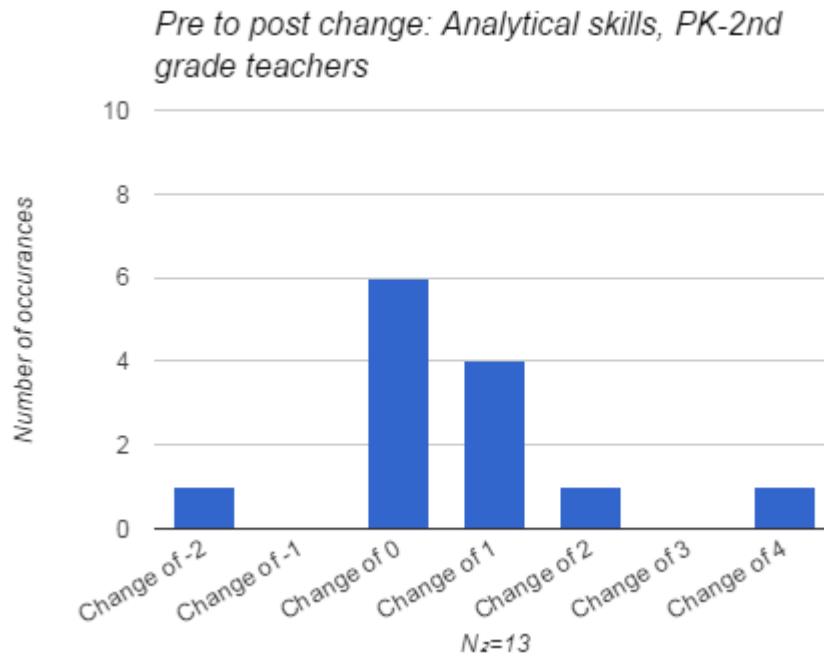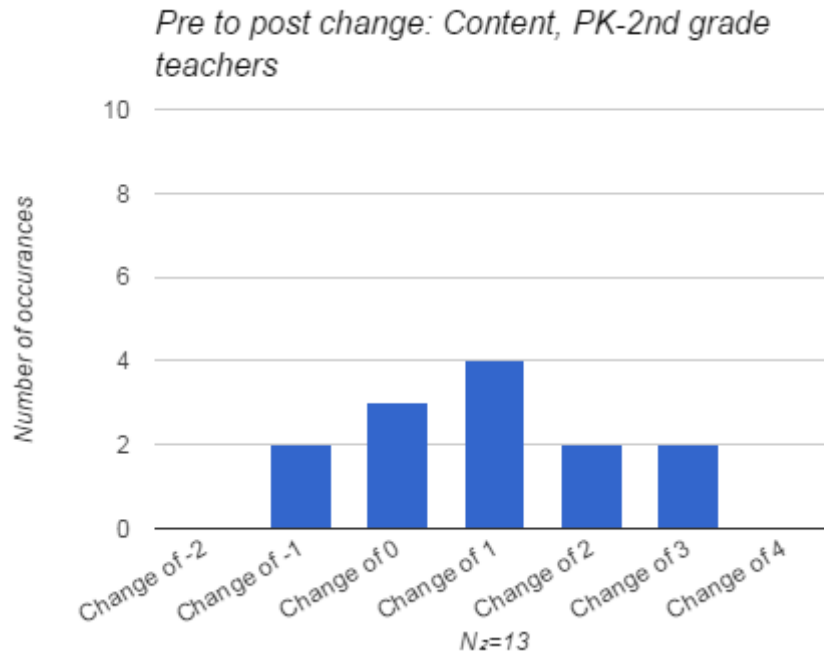**Appendix H: Mentions of "Vocabulary" in participant responses**

Percent of total responses that mentioned "Vocabulary" in pre-assessment				
	What did you need to know in terms of content?	What did you need to be able to do in terms of content?	What did you need to know in terms of language?	What did you need to be able to do in terms of language?
Change Over Time (N$_2$=19)	32%	21%	89%	42%
Energy and Matter (N$_2$=20)	35%	10%	80%	35%
Systems and Interactions (N$_2$=11)	9%	0%	73%	27%

Percent of total responses that mentioned "Vocabulary" in post-assessment				
	What did you need to know in terms of content?	What did you need to be able to do in terms of content?	What did you need to know in terms of language?	What did you need to be able to do in terms of language?
Change Over Time (N$_2$=19)	37%	11%	84%	26%
Energy and Matter (N$_2$=20)	25%	15%	45%	15%
Systems and Interactions (N$_2$=11)	9%	0%	73%	36%

Percent of total responses that <u>only</u> mentioned "Vocabulary" in pre-assessment				
	What did you need to know in terms of content?	What did you need to be able to do in terms of content?	What did you need to know in terms of language?	What did you need to be able to do in terms of language?
Change Over Time (N_2=19)	26%	11%	58%	16%
Energy and Matter (N_2=20)	30%	0%	55%	20%
Systems and Interactions (N_2=11)	9%	0%	45%	9%

Percent of total responses that <u>only</u> mentioned "Vocabulary" in post-assessment				
	What did you need to know in terms of content?	What did you need to be able to do in terms of content?	What did you need to know in terms of language?	What did you need to be able to do in terms of language?
Change Over Time (N_2=19)	11%	0%	26%	5%
Energy and Matter (N_2=20)	5%	0%	10%	0%
Systems and Interactions (N_2=11)	0%	0%	64%	9%

Appendix I: Grade level change

Pre to post change: Content, PK-2nd grade teachers
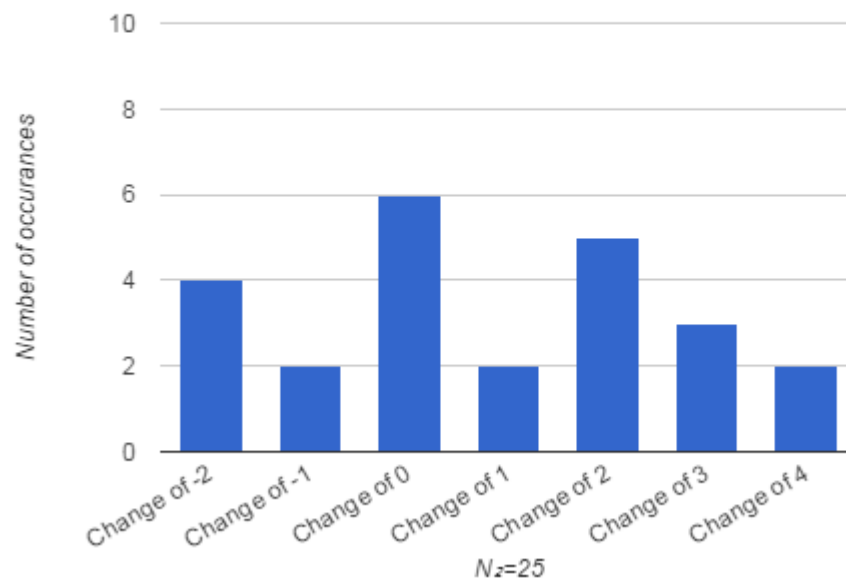

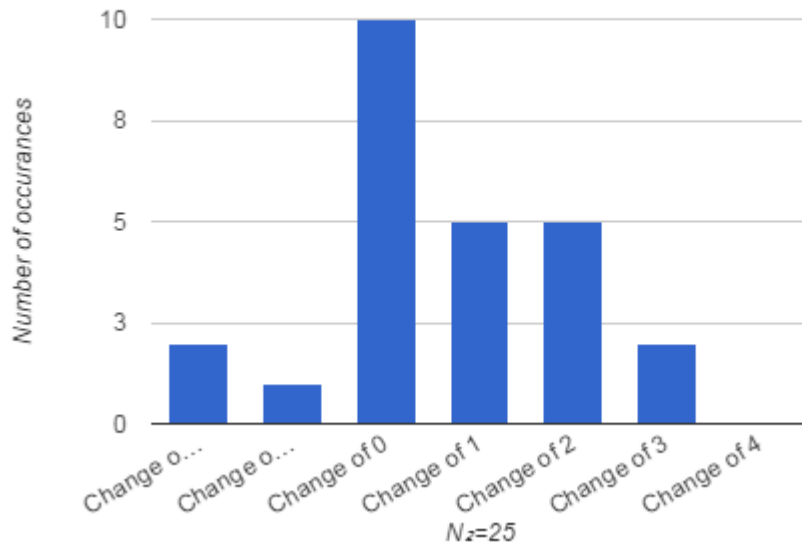Pre to post change: Analytical skills, PK-2nd grade teachers

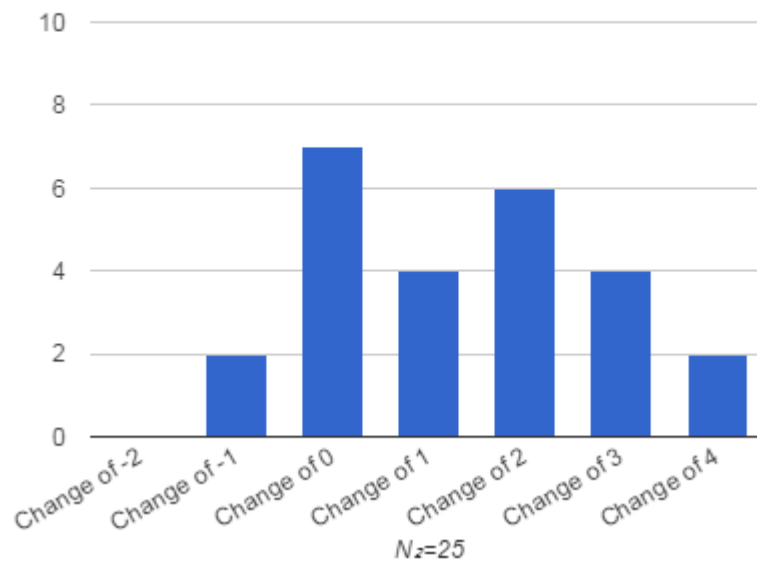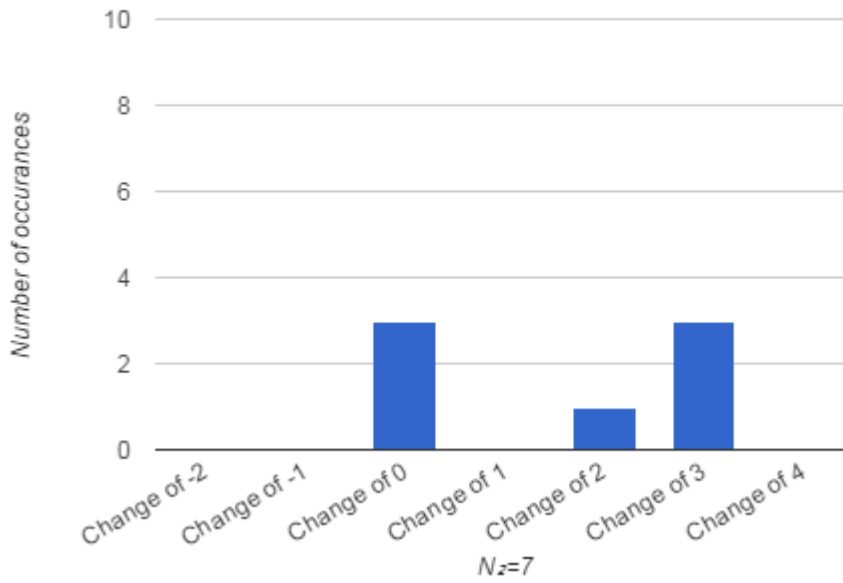*Pre to post change: Language, PK-2nd grade teachers*
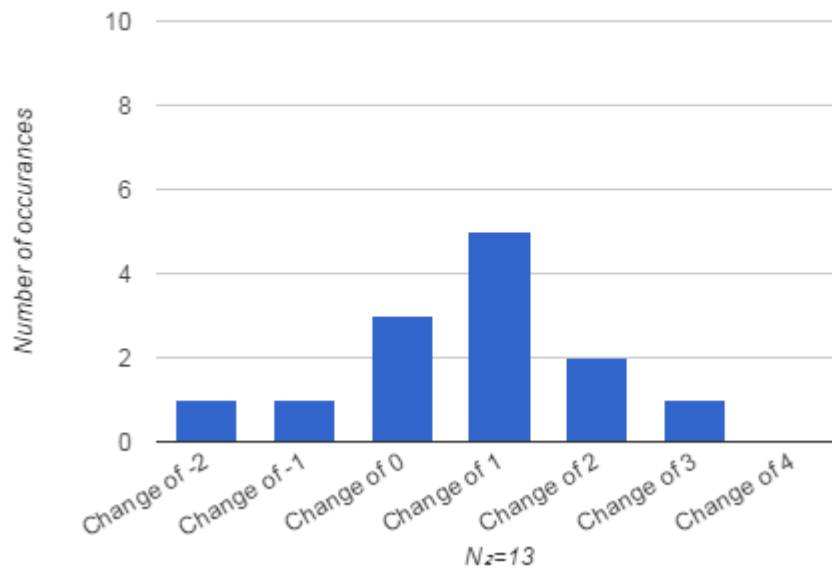
Pre to post change: Content, 3rd-5th grade teachers

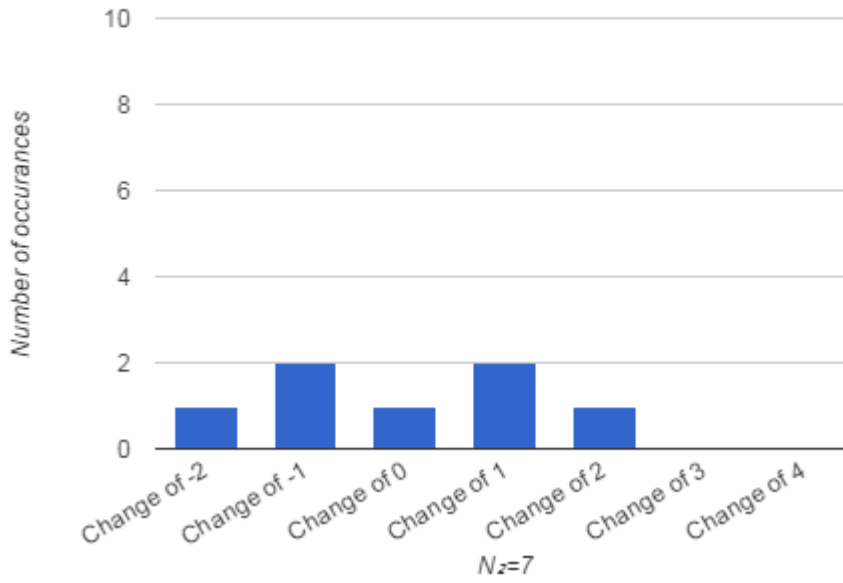*Pre to post change: Analytical skills, 3rd-5th grade teachers*

N_z=25

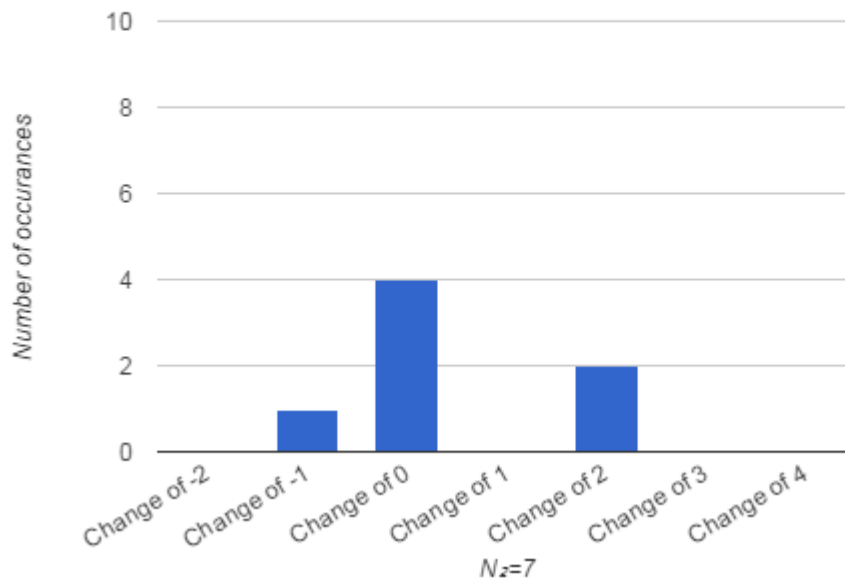*Pre to post change: Language, 3rd-5th grade teachers*

N_z=25

133

Pre to post change: Content, 6th-8th grade teachers

Number of occurances

Change of -2	Change of -1	Change of 0	Change of 1	Change of 2	Change of 3	Change of 4
0	0	3	0	1	3	0

$N_z=7$

Pre to post change: Analytical skills, 6th-8th grade teachers

Number of occurances

Change of -2	Change of -1	Change of 0	Change of 1	Change of 2	Change of 3	Change of 4
1	2	1	2	1	0	0

$N_z=7$

Pre to post change: Language, 6th-8th grade teachers

$N_z=7$

Appendix J: ELPA21 Language Functions and Forms Summary

A language **function** refers to what students do with language as they engage with content and interact with others. Functions represent the active use of language for a specific purpose. Students use language functions in order to express ideas, communicate with others, and show understanding of content in an academic setting.

> **In oral language** some common functions may include:
> - giving instructions
> - making requests
> - defending an argument
>
> **In academic writing** we use a range of specific functions in order to communicate ideas clearly. These include:
> - describing processes
> - comparing or contrasting things or ideas
> - classifying objects or ideas

Language **forms** deal with the internal grammatical structure of words and phrases as well as the word themselves. When one compares *boy* and *boys*, for example, or *man* and *men*, he or she is considering the relationship between different language **forms** or structures.

Language **forms** also include cross-curricular academic vocabulary - words or phrases frequently used across different content-areas. Cross-curricular academic vocabulary words typically describe or are related to academic *processes* and may include:
- verbs (e.g. *hypothesize, analyze*),
- complex prepositions, (e.g. *in contrast to, as well as*), and
- nouns (e.g. *comparison, conclusion, analysis*)

While functions address what we do with language, forms are the language structures and vocabulary that are used to support those functions.

Language learners need to acquire both the **functions** (uses/purposes) and the **forms** (structures + cross-curricular vocabulary) that make up the English language in order to reach higher levels of proficiency. Teachers also need to understand the language demands of a task as they relate to both **function** and **form** in order to best support students' language development.

Examples of Language Functions and Forms

Language Functions

- Expressing needs and likes
- Describing people, places, and things
- Describing spatial and temporal relations
- Describing actions
- Comprehending text or speech
- Retelling/relating past events
- Making claims
- Making predictions
- Asking Informational Questions
- Asking Clarifying Questions
- Expressing and Supporting Opinions
- Comparing
- Contrasting
- Summarizing

- Persuading
- Defending
- Analyzing
- Describing Cause and Effect
- Drawing Conclusions
- Defining
- Explaining
- Generalizing
- Evaluating
- Interpreting
- Sequencing
- Hypothesizing and speculating
- Summarizing

Language Forms

- Indirect/ direct object, subject/ verb agreement, pronouns
- Nouns, pronouns, adjectives
- Prepositional phrases
- Present progressive tense adverbs
- Past tense verbs, perfect aspect (present and past)
- Verbs: future tense, conditional mode
- Verbs and verb phrases in questions
- Questions with increasing specificity
- Sentence structure, modals (will, can, may, shall)
- Adjectives and conjunctions, comparatives, superlatives, adverbs
- Comparative adjectives
- Descriptive adjectives
- Increasingly complex sentences with increasingly specific academic vocabulary

- Verb phrases
- Sentence structure, specific vocabulary
- Verb forms
- Nouns, abstract nouns, pronouns, and adjectives
- Verb forms, indicative verb, declarative sentences, complex sentences, adverbs of manner
- Common, collective and abstract nouns, verb forms, nominalizations
- Complex sentences; increasing specificity of nouns, verbs, and adjectives; correlative conjunctions
- Language of propaganda, complex sentences, nominalizations
- Adverbs of time, relative clauses, subordinate conjunctions
- Modals (would, could, might), compound tenses (would have been)