Title

The Effect of an Integrated Science Program Designed for General Education Classrooms on the Academic Vocabulary of Fifth Grade Students with Special Needs and English Language Learners

Permalink

https://escholarship.org/uc/item/207618tw

Author

Laptes-Frangu, Mima S

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

The Effect of an Integrated Science Program Designed for General Education Classrooms on the Academic Vocabulary of Fifth Grade Students with Special Needs and English Language Learners

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Education

by

Mima S. Laptes-Frangu

June 2022

Dissertation Committee:
    Dr. Rollanda O'Connor, Chairperson
    Dr. Gregory Palardy
    Dr. Kinnari Atit

The Dissertation of Mima S. Laptes-Frangu is approved:

Committee Chairperson

University of California, Riverside

Acknowledgements

I am extremely grateful to my faculty advisor, Dr. Rollanda O'Connor, for the remarkable help and support received along these years. Your insightful advice and encouragement drove me to become a better researcher and writer during this long and arduous journey. Moreover, your high standards pushed me to give the best of my work to accomplish this extraordinary task.

I would like to extend my sincere thanks to the faculty members of my committee, who were so supportive in finalizing my work. Special thanks go to the remarkable faculty and staff members in the Graduate School of Education. Having such a wonderful big team backing me up made my journey so much more interesting and fruitful. From among these great people, I would like to acknowledge Heather Killeen for her continuous assistance with my questions and requests for help.

I want to acknowledge my colleagues who were there when I needed to work on the inter-rater reliability in the field – thank you for your time and effort, Brian! – and who gave me advice when needed, as true friends besides being wonderful colleagues. I want to mention here Olivia, Emad, Joyce, Sounghwa (not in a specific order) – thank you so much for being there for me!

This endeavor would not have been possible without the extraordinary help and support that Dr. Jodye Selco extended to me from the start to the end of this demanding project. Your mentoring was the guidance that I needed through the challenges endured, and for that I am forever grateful! Many thanks to Juanita Chan for her huge help with this study!

I am also deeply thankful for being a part of the team of researchers in Dr. O'Connor's CHAAOS lab. I want to acknowledge especially Dr. Kathleen Bocian, from whom I have learned so much during my time there. Teaching me tricks of the trade was a great way of learning how to conduct research.

I could not have been successful in this journey without my wonderful family. Deepest appreciation to my incredible children, Andrei and Raluca, who were there for me during these trying times, and never left my side. You knew when to step in and make my everyday life easier so I could have more time to do my research and work on my dissertation. Knowing that I am doing this for you gave me strength to accomplish this enormous effort, and I thank you for that! Deepest gratitude to my dear parents, Cincinella and Valentin: you made this happen, by bringing me into this world and raising me so well, for which I thank you both deeply! Many thanks to my husband Florin for his help during this program, despite the hard times.

Dedication

To my beloved children, Andrei and Raluca, who were my rocks during this journey: you can achieve anything you dream of, in spite of the challenges in your way.

"*Our greatest glory is not in never failing, but in rising every time we fail.*" – Confucius

ABSTRACT OF THE DISSERTATION

The Effect of an Integrated Science Program Designed for General Education Classrooms on the Academic Vocabulary of 5th Grade Students with Special Needs and English Language Learners

by

Mima S. Laptes-Frangu

Doctor of Philosophy, Graduate Program in Education
University of California, Riverside, June 2022
Dr. Rollanda E. O'Connor, Chairperson

The Next Generation Science Standards (NGSS) have high requirements regarding science knowledge and academic vocabulary to be taught to students using research-based curriculum and teaching techniques. Students who receive Special Education services or who are English Language Learners (ELL) are known to encounter difficulties learning new and complex scientific concepts; however, according to NGSS, they are accountable to the same standards as their typical peers. Numerous studies show that Project-Based Learning (P-BL), together with scaffolding and group learning, have been used successfully in science classrooms at different grade levels (Filippatou & Kaldi, 2010; Simons & Klein, 2007). Previous research also supports integration of P-BL with academic literacy to develop a strong science vocabulary (Santau, Maerten-Rivera, & Huggins, 2011). Since school districts started to implement NGSS (2015-2016), research conducted to indicate how the new requirements and techniques to teach them address the various needs of a diverse student body has been scarce, and there is still need for more studies that assess different ways in which the new science standards are being applied in classrooms. This study searches for evidence of whether a new science curriculum (i.e., The Common Labs, or CL) that is used currently in a school district in Southern California and which was designed as a P-BL, integrated science curriculum for General Education classes, is also effective for students who receive Special Education services and/or who are ELL. Vocabulary gains of students who received the CL or business-as-usual (BAU) curricula were compared before and after engaging in a specific science unit.

Due to the pandemic shut down of all schools, the study suffered large attrition in posttest scores that were used as the measure for students' science vocabulary improvement at the end of the science module. Thus, imputation was used to conduct the data analyses using the entire sample of pretests administered to students prior to the pandemic restrictions. The results showed that students in BAU classes performed better than their peers in CL classes, as measured by the posttest scores of the science vocabulary multiple choice test. Observation data revealed few instructional differences between the two conditions.

Table of Contents

List of Tables

List of Figures

Introduction

Vocabulary knowledge plays a fundamental role in learning outcomes across all subject areas. Hence, vocabulary deficiencies have been known to be tied to poor student outcomes. In the most recent years, the number of students receiving special education services in K-12 school system was 6.5 million, representing approximately 13 percent of the public schools' student body (National Center for Education Statistics, 2016). The percentage of English Language Learners (ELL) in public schools was 9.3 percent in 2013-2014, following an increasing trend over the past few years (NCES, 2016). In accordance with the increasing number of ELL and students with special needs (who are receiving special education services), and likely being affected by it, the reading performance of middle school students was lower in 2015 than in 2013, according to NCES (2016). Middle school reading skills are highly influenced by the reading knowledge achieved in higher elementary grades. Therefore, fifth grade reading comprehension across the curriculum (including science as well as mathematics, social sciences, and all the fields of the curriculum) becomes a pivotal instructional area to prevent students' academic decline in multiple subject matters in middle and high school, their dropping out of school, and all the consequences that derive from that (Connor, Alberto, Compton, & O'Connor, 2014).

For students who are ELL, learning science content may be especially difficult because science concepts tend to be isolated from other curriculum subjects such as English literacy (particularly reading comprehension) and mathematics (Cuevas, Lee, Hart, & Deaktor, 2005; Lee, Maerten-Rivera, Penfield, LeRoy, & Secada, 2008). Science researchers (Meyer & Crawford, 2015) have suggested that science could become more contextualized and easier to be internalized by students if science vocabulary and activities were integrated with other subject matters such as English language arts and mathematics. There were also studies and reports showing that students' performance on standardized tests in science along with reading and mathematics were also improved when the concepts were contextualized and presented in a meaningful an interesting way to the students (Cohen, 1990; Ghosh, Hokom, Hunt, Magdaleno, & Su, 2008). The internalization process can be described by the integration of the new concepts learned with different other supporting concepts, whether they are previously known or acquired at the same time with the newly learned ones, then moving the newly-acquired concepts into long-term memory. In the next section, I am describing research that has generated evidence that the integration of science with other subject matter strategies cultivates improved outcomes in science for general education students, as well as for ELL and students with special needs (Hmelo-Silver, Duncan, & Chinn, 2007; Meyer & Crawford, 2015; Pine et al., 2006; Simons & Klein, 2007).

Addressing academic vocabulary in science with high precision, especially for ELL and students with special needs, is paramount to the comprehension of science concepts (Bravo & Cervetti, 2014). Without the most suitable support from the teachers, the linguistic exigencies of science can impede ELL's and students with special needs' comprehension of already difficult scientific concepts. Several recent studies focused on the gains of students in science classrooms when responsive learning contexts and scientific inquiry are used to improve students' science academic vocabulary (August, Branum-Martin, Cardenas-Hagan, & Francis, 2009; Bravo, Cervetti, Hiebert, & Pearson, 2007; Bravo & Garcia, 2004).

In this study, I am observing, describing, and analyzing the results of a new, hands-on/project-based, integrated and applied science curriculum named The Common

Laboratories (The Common Labs or CL), that was implemented in an elementary school district in the area. The elements that this program is based on are: Project-Based Learning (P-BL) through group work, scaffolding, integration of science with language arts and mathematics, and use of technology as a classroom aid. This combination of elements is intended to improve students' academic vocabulary and science knowledge. The Common Labs program was designed for general education classrooms, but it is applied also to students who are ELL and receiving Special Education services who are mainstreamed. My intention was to study its impact on these English Language Learners (ELL) and students with special needs.

Hence, the present study seeks to answer the following research questions in order to present a scholarly evaluation of the Common Labs program:

1. Do students with special needs (who receive special education services and have mild/moderate disabilities) and students who are ELL who are exposed to the Common Labs program improve academic vocabulary and general science knowledge more than the matched group of students who are not exposed to this curriculum, after an 8-to-10-week period of instruction, as shown by their test results?

2. Is the effect of participation in the Common Labs program similar for students with special needs, ELL, and their typical peers, as indicated by the test results?

In this dissertation, I am addressing the theoretical foundation of The Common Labs curriculum and how the theoretical elements are intertwined, which may improve outcomes for the students with special needs and ELL. By reviewing the literature describing these instructional techniques in several previous studies, I intended to reveal research support for the instructional elements of The Common Labs curriculum. Following the literature review, I am presenting the research participants, methods, and analytical tools that the study design encompasses, in order to test the effect of The Common Labs for students with special needs and/or who are ELL.

Methods of Teaching Science in Elementary Schools

Using traditional, lecture-type instructional techniques in science classrooms is becoming outdated in an increasing number of school settings (Dalacosta, Kamariotaki, Palyvos, & Spyrellis, 2009). Instead, a rising number of science teachers use different types of interactive, more engaging methods in their classroom as aids to the traditional teaching methods. Subsequently, they may be obtaining better results (Kawalkar & Vijapurkar, 2013; Mathan & Koedinger, 2005); however, the effect of particular features of curricula have rarely been compared for students with special needs, including students with disabilities or who are ELL. Students differ in their ways of learning. Brewer (2004) discusses the ways in which the learning process can be multi-sensorial, collaborative, and practical/experimental. According to Brewer, there are different ways of approaching the instructional process, both in traditional instruction and in a project-based approach. Observing these different approaches gives us an understanding of the impact of hands-on-based instruction (or project-based) *per se,* as well as when hands-on instruction is supported by group work and scaffolding.

When working within a P-BL framework, it has been found that specific teaching methods are needed to support P-BL in being successful. Some of these methods are scaffolding (Simons & Klein, 2007) and collaboration through group work (Filippatou & Kaldi, 2010). These instructional features will be important to document in my study in both conditions (CL and BAU – *business-as-usual* classes).

Furthermore, by integrating science into everyday life, teachers can offer students numerous opportunities to interrelate with course ideas. This integration may be not only beneficial (Saye & Brush, 2001), but also induce a change from the usually competitive classroom environment (through the focus on grades and individualized work and answers), to one that is more collaborative. The collaborative learning environment is more student-centered and more focused on students' knowledge improvement than on grades, according to the view of learning as being multisensorial, experiential, and interactive (Brewer, 2004). Student-centered instructional activities are intended to provide students with a plethora of opportunities to take an active role in their own learning process by transferring the responsibilities of organizing, investigating, synthesizing and, in the end, assessing the content, from teachers to students. Implementing a collaborative classroom environment gives the students the chance to learn collaboration skills that will be beneficial to them later (Saye & Brush, 2001).

The integration of different elements of the curriculum design such as project-based instruction and applied and collaborative learning may be beneficial for the learning process when working with different types of learners, from typically-developing to students with special needs and second-language learners. Project-based instruction has been shown to be successful in general education classrooms, as well as with ELL and students with special needs (Bell, 2010; Filippatou & Kaldi, 2010). Similar research showed that integrating science topics with other curriculum subjects such as reading or mathematics could also be helpful to students' understanding of difficult science concepts (Brophy, Klein, Portsmore, & Rogers, 2008; Vadasy, Sanders, & Nelson, 2015).

Project-Based Learning in Science Classrooms

Learning science in elementary grades has been difficult even for general education students, particularly in a traditional, lecture-based classroom (Amos & Boohan, 2003). Science learning for students with special needs and ELL has been even more challenging than for their general education peers due to lower levels of vocabulary and reading comprehension skills to grasp complex concepts taught in science classrooms. To reform science education, a constructivist framework was believed to be the most suitable due to its view of learning as being an interpretative process that involves the construction of individual knowledge through social collaboration (Tobin, Briscoe, & Holman, 1990). Constructing strong knowledge in the elementary grades may be enhanced through a project-based approach and directly involving the students in the instruction process (Erdogan & Campbell, 2008). Project-based learning (P-BL) is specifically successful in science classrooms, where the hands-on experiences cannot be substituted by teacher's lectures exclusively (Erdogan & Campbell, 2008; Tobin et al., 1990).

In her review of recent research on P-BL, Bell (2010) collected the most relevant studies and touched on the main benefits that this type of instruction brings to students. Students learn how to increase their self-reliance through carefully planning their approach to the problem in collaboration with peers. The enhancement of students' collaboration skills is gained through social learning that is specific to a P-BL environment. Furthermore, embedding the newest technology, such as computer- and/or internet-based learning, or use of smart boards, in the P-BL framework has also been shown to improve students' achievement of science skills. All these characteristics make project-based instruction beneficial for students (Bell, 2010).

To delve further into potential P-BL advantages when compared to traditional classroom instruction, Filippatou and Kaldi (2010) conducted a study with 94 fourth-graders in Greece, including 24 students with learning disabilities. The authors observed the students during a period of 8 weeks of carefully-planned classroom activities that were implemented for 2-3 hours per week. The curriculum reviewed during the study was about sea animals and included their classification, reproduction, their food chain, how sea animals are part of human nutrition, and which sea animals are threatened or almost extinct. The specific instructional module was taught using the educational theory of P-BL, with activities that included field-based visits, hands-on activities, experts-lead talks in classrooms, as well as technology-based sources such as DVDs, pictures, small brochures, and game-based learning. Their results showed that the students with disabilities who participated in this program scored higher on the posttest scores regarding knowledge of the topics in the module taught using P-BL approach than before this approach was applied. They also developed a greater sense of teamwork and stated that they found the P-BL experience more enjoyable than the previous traditional classroom teaching (Filippatou & Kaldi, 2010).

To reveal the effectiveness of P-BL instruction in regard to students' background, Cuevas et al. (2005) investigated teachers' effectiveness in intertwining science and literacy with students' previous knowledge. Their study was conducted with 25 third and fourth-grade students with different levels of academic achievement and mastery of English language. The students were presented a problem regarding the impact surface areas have on evaporation rate. The students were asked to use inquiry-based investigation in their search for the answers. The teachers were asked to help students' investigation by asking a series of questions from the protocol developed by the research team. Next, the students were asked to extend their findings to water evaporation from oceans, lakes, and rivers. At the end of the school year, Cuevas et al. found that the low-achieving, low-SES students made remarkable gains compared to their high-achieving, middle-SES peers in their inquiry skills improvement, measured pre- and post-intervention. The authors concluded that problem-based and collaborative learning environments can foster science inquiry for all elementary students, regardless of their culture and/or native language.

Moreover, additional research has studied how P-BL impacts learning through inquiry and what skills it fosters. In her review of the literature in the field, Bell (2010) found that students learning within a P-BL environment learn more easily and become self-sufficient through the group dynamics. Working in small groups also helps students improve or start developing their interdependency and collaborative skills out of consideration for the greater good. Bell also emphasized how these skills led to superior intrinsic motivation that emerged from students' accountability toward their peers, which had greater impact than the accountability toward the teacher. Bell proposed that social skills such as communication, collaboration, and negotiation are not only being developed and/or improved within group work, but may also lead to more mature young people in the future, who will be self-confident in using their creativity fostered by their collaborative abilities, productive communication, and fruitful teamwork.

In a study conducted by Kuo, Hwang, and Lee in 2012 with a sample of 58 fifth-grade students, the authors proposed a hybrid approach to increasing the problem-solving ability of students. The researchers used various instructional strategies with the two groups of participant students, in 4 different phases. The main methods were the collaborative problem-based learning (P-BL) in the treatment group, compared with mostly the modeling approach for the control group. In Kuo et al.'s (2012) study, within the treatment group, students learned by intertwining the demonstrational instruction model with the collaborative learning method; the specific instructional methods used were also supported by coaching and scaffolding. Meanwhile, the control group used only the modeling approach, also supported by coaching and scaffolding. There were also other instructional techniques used in the treatment group, such as group versus individual learning, followed then by learning without coaching and/or scaffolding. Similarly, within the control group these teaching methods were also used with modeling, but not together with collaborative learning. The results showed that the performance of the experimental group that used collaborative problem-solving instruction was noticeably higher than within the control group that used exclusively modeling instruction, even though both groups had coaching and scaffolding added to the instruction, then taken away later on. The results were higher even for the students with low to middle academic achievement before the study, who may be similar to the students in the present study.

Group Work in Science Classrooms

Among the instructional conditions utilized by science teachers, student collaboration, either in small groups or in larger groups, may be an essential one. During the science instruction time, students in CL are supposed to closely collaborate with members of their groups in order to develop and execute a plan to accomplish the required task, which has been supported by several research teams (Filippatou & Kaldi, 2010; York-Barr, Ghere, & Sommerness, 2007). When teachers encouraged group work among the students with learning disabilities (York-Barr et al., 2007), students' behavior went from a passive role to becoming involved in the groups' work and obtaining results, and subsequently gained self-efficacy beliefs. Several other studies focusing on improving literacy skills of students with either special needs or ELL also uphold the importance of students either working in small groups or pairs on acquiring academic vocabulary knowledge (Baker et al., 2014; Graves, Brandon, Duesbery, McIntosh, & Pyle, 2011; Lawrence, Crosson, Paré-Blagoev, & Snow, 2015; Lesaux, Kieffer, Kelley, & Harris, 2014). Within the wider framework of P-BL, group work and the collaborative relationship that develops among the members of those groups may become a strong foundation for acquiring new academic knowledge (Cuevas et al., 2005; Silverman, Proctor, Harring, Doyle, Mitchell, & Meyer, 2013). Students who are ELL have been found to benefit the most from being placed for group work in heterogenous groups that contained students with various levels of language proficiency and academic skills, whether the instruction is provided directly by the teacher or the students are collaborating within the group on their classwork (Cohen, 1990; Foorman & Torgesen, 2001; Ghosh, et al., 2008).

Among several studies conducted on group work's effects on academic achievement of students who are ELL, Amaral, Garrison, and Klentschy (2002) performed a study that summarized the results of a four-year science educational program with elementary students who were ELL. The student sample consisted of 615 fourth grade students and 635 sixth grade students. Of the participant students, approximately half were ELL (293 in fourth grade and 338 in sixth grade) with different degrees of English language proficiency. Students were grouped by their English proficiency and each of the two major groups (limited English proficiency and English proficient) were further grouped into subgroups based on students' English proficiency level. Besides the small group work, other teaching strategies used in the classrooms were independent work and whole-class activities where appropriate. From these methods, small group work, also known in the literature as cooperative learning, is believed to help students develop their social/collective skills and strengthen academic vocabulary through opportunities for discussing with and learning from their peers. In their small groups, students who were ELL were able to partake in the classroom activities without the unease of speaking in front of the entire class, not knowing the right answer or not being able to offer an adequate explanation in a language that they are still learning (Amaral et al., 2002). The authors concluded that the academic achievement of students who were ELL improved proportionally with the number of years they were in the program, both in science and literacy skills.

Group work was found to be an important structure in a more recent study conducted by Filippatou and Kaldi (2010). As previously described, their study showed that group work was beneficial to elementary school students with special needs (such as students who receive Special Education or who are ELL) and that it worked well in conjunction with P-BL. Their first study design used a pretest and a posttest which showed how group work improved learning outcomes, the students' engagement in the instructional process, and easier acceptance by their peers within the group work, all of these being major outcomes that students may acquire when group work is used extensively as an instructional technique (Filippatou & Kaldi, 2010). This study showed the importance of group work in particular among other instructional methods.

A thorough review of various elementary engineering curricula and subsequent activities was done by Brophy et al. (2008). In their review, the authors surveyed several STEM-based instructional programs in different parts of the country. At the elementary level, the main programs examined were *Engineering in Elementary* (EiE), a program primarily funded by the National Science Foundation (NSF), and LEGO Engineering, a program operated by the Tufts Center for Engineering and Educational Outreach (CEEO). Both programs are designed within a P-BL frame. EiE has been shown to improve students' science understanding through interconnecting it with engineering and technology concepts in a hands-on, inquiry-driven instructional environment (Brophy et al., 2008). Furthermore, LEGO Engineering, besides confirming a general improvement in students' aforementioned skills, has been shown to distill and bring forward teachers' own science-teaching skills. Teachers' strength in teaching a high-tech, scientific-based curriculum was one of the major influencers in students' academic outcome. The highly-dense STEM-based instructional programs using a P-BL framework showed students' improvement of their science/engineering skills, which Brophy et al. attribute to using a hands-on approach. Furthermore, the state standards require all students to be held accountable by having their achievement graded by the same standards, regardless of classification as ELL and having special needs. As additional research has indicated, students being in control of their own learning environment is beneficial, helping them improve their analytical and problem-solving skills (Harskamp & Suhre, 2006; 2007).

Teacher Scaffolding in Science Classrooms

Although the constructivist theory was Piaget's notion, Vygotsky had a paramount contribution to constructivism through his Socio-Cultural Theory (SCT), which explains building knowledge based on socio-cultural interactions between the learner/student and adults or peers with a higher ability level than the learner (Stone, 1989; Vygotsky, 1978). To explain his learning theory further, Vygotsky developed the concept of the Zone of Proximal Development (ZPD), which is the foundation of modern instruction (Foorman & Torgesen, 2001). The Zone of Proximal Development was described as "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance, or in collaboration with more capable peers" (Vygotsky, 1978, p. 86). With ZPD comes into play scaffolded instruction, which was developed as a result of, and became synonymous with ZPD. Since Vygotsky's ZPD description in the early 20th century, followed by its usage in education in the early 1960's, scaffolding has become the applied representation of the ZPD concept. Teacher scaffolding can be productive for students when used appropriately.

As Simons and Klein (2007) observed on a sample of 111 low-performing seventh-graders taking a STEM course, scaffolding can supplement teachers' various other techniques, while not replacing them entirely. One of the experimental groups of students received scaffolding during the entire 9-week period. Scaffolded students performed better than the control group who did not receive any scaffolding, and slightly better than the second treatment group who received scaffolding only optionally. At the posttest, the lowest-scoring students in both experimental groups performed only slightly above the level of the control group. The authors' conclusion was that students' attitude towards the scaffolded instruction as well as their prior knowledge in the field were more influential than the teaching strategy itself. The students who scored the highest at posttest in both treatment groups were the ones who performed at the highest level during the intervention. Similarly, Pine et al. (2006) found that scaffolding intertwined with instructional technology was more helpful than mere usage of technology in classroom. Other researchers also revealed that the degree of help provided to students should depend on those students' needs. As an example, programs that deliver early feedback to the students who need more help, might improve their metacognitive processes of detecting, rectifying, and subsequently learning from errors (Mathan & Koedinger, 2005).

Recent research revealed the importance of instructional scaffolds both in reading and science instruction (Meyer & Crawford, 2015). Scaffolding facilitates students' understanding of how scientific concepts can be applied to real life and what they encompass by providing the most suitable support to students. With adequate support, students can attain proficiency in the targeted area, which would otherwise be beyond their unassisted endeavors (Vygotsky, 1962; Vygotsky, 1978; Wood, Bruner, & Ross, 1976). The scaffolding process is one of the most effective supportive teaching methods in education today (Hmelo-Silver, et al., 2007; Reiser, Tabak, Sandoval, Smith, Steinmuller, & Leone, 2001; Simons & Klein, 2007), and has been utilized since Vygotsky's Zone of Proximal Development was first introduced (Palincsar, 1986). Although August et al. (2014) recommend scaffolding as an effective technique with ELL, they note that scaffolding learning of students who are ELL may require more time than with their typical peers to accommodate task specific vocabulary in a second language.

Scaffolding methods used in combination with Project-Based Learning (P-BL) have been found to foster the improvement of students' conceptual understanding of science (August et al., 2009, Lee et al., 2008). Furthermore, P-BL has been shown to have a positive effect on students' learning as well as on teachers' use of various teaching strategies, such as student-centered instruction, enhanced collaboration within classrooms, scaffolding, or use of appropriate technology in classroom (Ertmer &

Simons, 2006; Saye & Brush, 2002). From early childhood through adulthood, P-BL supports skills development while also supporting learners' natural impulse to explore different things and phenomena. Several more studies on P-BL suggested that this type of instruction is effective, especially in fields such as science and mathematics (Bell, 2010; Filippatou & Kaldi, 2010). Within the Common Labs program that I have studied, scaffolding was an important element of P-BL and was the leading strategy employed by teachers.

Improving Academic Vocabulary in Science Classroom

Although curriculum plays a fundamental role in science classrooms, science content also relies on adequate vocabulary knowledge. Focus on particular vocabulary words may be equally important in achieving a successful outcome for vocabulary improvement and subsequently for comprehending science concepts. Previous research has shown that teaching academic vocabulary words and providing opportunities to use them in small group discussion, improves reading comprehension skills and overall language development not only for typical students and ELL, but also for students with special needs (Apthorp, Randel, Cherasaro, Clark, McKeown, & Beck, 2012; Graves et al., 2011; Nelson, Vadasy, & Sanders, 2011; Vadasy & Sanders, 2010). General academic and multiple-meaning words are important for students to understand the new concepts, and are generally words with several connotations, used across differing curricula (Beck, McKeown, & Omanson, 1987; Biemiller, 2010; Sibold, 2011).

Recent research (O'Connor, Gutierrez, Teague, Checca, Kim, & Ho, 2013; Silva & Cain, 2014; Vadasy et al., 2015) has described the third grade as being a fundamental benchmark for student success in school and later in life. If a child has not yet developed skills in basic reading comprehension by third grade, we can expect s/he will struggle in later grades, lose interest in school, possibly drop out, and in many cases, live through major negative consequences in their lives. Hence the importance of developing vocabulary for reading comprehension in the intermediary/higher elementary grades. Integrating reading comprehension across subject matters could benefit students by allowing the reading skills acquired in the English Language Arts (ELA) classes to be practiced further in other classes, such as science. More studies on the efficacy of different interventions directed on improving the science and academic vocabulary achievement for students who are ELL were conducted more recently. These showed that P-BL programs blended tightly with academic literacy were the most successful in helping students who are ELL (and by extrapolation, students with special needs) in their developing of a solid academic vocabulary (August et al., 2009; Santau, Maerten-Rivera, & Huggins, 2011).

Fifth grade is generally known to be a difficult grade, considering the amount of new concepts taught and thus requiring subsequent high-leap academic gains. Therefore, providing vocabulary-building opportunities integrated throughout the curriculum areas may be important (Reardon & Galindo, 2009). Current research in the field of science teaching and reading comprehension skills validates the widely accepted concept that higher-quality environment and instruction result in accelerated learning (Manset-Williamson & Nelson, 2005).

Reviewing previous research regarding the academic vocabulary in science classes, Lee and Fradd (1998) emphasized the difference between knowing and doing science, while acknowledging the importance of talking science, hence the use of group work and P-BL while building academic vocabulary. Teachers also have a vital role in the instructional congruence between "the nature of science and the language and cultural experiences of the students" (Lee & Fradd, 1998, p. 18). While their literature review is focused on students from non-English-language backgrounds, the authors emphasize the possibility of using similar approaches to obtain comparable results for various other student groups and even different subject matters.

In view of the inquiry process as being fluid and growing in complexity as students (together with teachers) evolve in their gaining of both science and inquiry skills, P-BL instruction could help ELL overcome their linguistically-driven lack of understanding in learning new scientific concepts through the extensive use of academic vocabulary (Fradd, Lee, Sutman, & Saxton, 2002). Fradd et al. conducted a study with approximately 900 fourth graders per year for 3 years. Students were attending 7 inner-city and suburban elementary schools; among the participants, the majority were ELL. The authors examined the interconnection among language/literacy and science, as they were used in science activities with ELL. Students who are ELL traditionally achieve less well in science than native English speakers, as illustrated by consistently lower scores in the National Assessment of Educational Progress (NAEP) Technology and Engineering Literacy (TEL) tests several years in a row (NAEP, 2014, 2015; NCES, 2016, 2017). Fradd et al.'s program named Science for All (SFA) emphasized P-BL science instruction for elementary school ELL students, and incorporated small group discussion to foster vocabulary acquisition. Their findings lead toward recommending a wider and tighter collaboration among teachers, researchers, scientists, and also policy makers, to guarantee that ELL attain science knowledge at the same level as their native-speaking peers, as expected by the Next Generation Science Standards (NGSS).

To conclude, the structured P-BL that the CL program employs is thought to deliver positive outcomes for students when is tightly intertwined with small working group settings, which encourage the initiation of, and utilization of students' prior knowledge and science vocabulary. This process occurs more often in small working groups through the mutual encouragement of active participation of all group members (Filippatou & Kaldi, 2010; Schmidt, Rotgans, & Yew, 2011; York-Barr et al., 2007). Small group work has also brought benefits to students by encouraging both students and teachers to have greater focus on the tasks at hand (Lawrence et al., 2015; Lesaux et al., 2014; Silverman et al., 2013). The fluid scaffolding that is embedded within the CL program (as discussed above) is also building on collaborative group work by facilitating teacher's use of scaffolding with a smaller number of students, which may contribute to the effectiveness of the Common Labs instruction. All these instructional features together form a robust foundation for potential improved learning and acquisition of vocabulary knowledge of students in the CL program that is presented in this study.

Purpose of This Study

The goal of this study is to examine The Common Laboratories program that is running in the participating school district by conducting a quasi-experimental design with matched participants. I compared class outcomes in six schools that were participating in the program with a matched comparison group from six other schools in the same district that were engaged in BAU science instruction. The Common Labs curriculum was set up as Project-Based Learning (P-BL) due to ample research suggesting this type of instruction may be effective (see Bell, 2010; Filippatou & Kaldi, 2010). In particular, the present study has analyzed the science curriculum (the Common Laboratories) used by the classroom teachers as it affected academic outcomes of students with special needs and/or students who were ELL's development of academic vocabulary and science knowledge.

As illustrated in this review of the literature, students' improvement in science learning might occur when learning is based on a well-structured P-BL program and supported by other teaching methods such as scaffolding (Simons & Klein, 2007), group work (Filippatou & Kaldi, 2010), and hands-on science instruction (Pine et al., 2006). Furthermore, the integration of the science program with other subject matters across the curricula (such as English language development) could enhance the ways through which ELL and students with special needs are exposed to science context. In the present study, science is being integrated with mathematics and English literacy, which was observed during the field stage of the study using the Classroom Observation Tool.

The Common Labs program was created to integrate literacy skills such as listening, speaking, reading and writing following the ELA Common Core State Standards' framework with science learning, through employing close and critical reading assignments and accountable talking tasks. Compared to the old state standards, where the students had to *know* a concept, the newly state-adopted science standards – Next Generation Science Standards (NGSS; California Department of Education, 2018) – have students learn science by *doing* it. In order to be able to efficiently accomplish this hands-on type of learning, students also need to be able to comprehend what they read about it first; hence, the integration of the vocabulary skills into the science curriculum. The NGSS were approved by the California State Board of Education (SBE) in September 2013 (California Department of Education, 2017), but the school districts have started implementing them a few years after that, during the Transition phase (2015-2016) identified and anticipated by the California Department of Education (California Department of Education, 2017). The body of literature in the sub-field of elementary science education following the implementation of the new NGSS is still growing and there is a high need of more studies that are assessing different ways in which these state standards are being applied at the school and classroom level.

For the purpose of this study, I chose a science module that is being taught to fifth graders. According to the new NGSS for fifth grade, specifically standard 5-LS1-1 – *Support an argument that plants get the materials they need for growth chiefly from air and water* – the students will have to demonstrate how the matter and energy flows are organized in plants, as well as what interactions and dynamics occur to transport matter into, out of, and within the living systems. Compared to traditional teaching methods and previous state standards, The Common Labs program seeks to improve students' science skills and academic vocabulary through integrated-across-curriculum, hands-on science instruction that follows the NGSS.

The Common Labs program has been designed to improve multiple skills across the subject matters, including mathematics, vocabulary, and reading comprehension, using academic conversations within classrooms together with science-specific instruction. Cognitively-challenging questions and academic language scaffolds are also used to help students develop wider cognitive skills (Meyer, Coyle, Halbach, Schuck, & Ting, 2015; Nagy & Townsend, 2012). By improving students' academic vocabulary skills through the employment of group work and problem-based learning discussion, it is expected that the Common Labs program could facilitate the full implementation of the NGSS with students who have special learning needs (special education and ELL).

Methods

The present study is analyzing the effects of an integrated, hands-on science program that was designed for general education classrooms, on the academic vocabulary and science learning of fifth grade students who are English Language Learners (ELL) and students receiving Special Education services, when compared with their typical peers and with students in similar categories in control classrooms.

Participants and Setting

The students of participating science teachers are enrolled in 6 schools within an urban school district located in Southern California (XUSD), in general education classrooms that also include English Language Learners (ELL). The schools are all Title I schools, with families having low socio-economic status (SES) and a high percentage of ELL students, most of whom are of Latino origins. XUSD's student body demographics are as follows: from a total of 25,684 students enrolled, 21,449 (83.51%) are of Hispanic background; 21,195 (82.52%) are receiving free or reduced lunch (low SES); 7,082 (27.57%) are ELLs, and 2,732 (10.64%) are students with special needs who receive special education services. This information is detailed in Table 1 (California School Dashboard, 2018):

Table 1
Unified School District (XUSD) student demographics. California School Dashboard, 2018

Enrollment by Ethnicity

USD	Asian	African-American	Hispanic/ Latino	White	Other	**Total**	Free & red. lunch	ELL	SpEd
Number of Students	222	2,541	21,449	910	562	**25,684**	21,195	7,082	2,732
Percent	0.86	9.89	83.51	3.54	2.19	**100.00**	82.52	27.57	10.64

Six 5th grade teachers and their students in science classes from 6 schools in the school district who participated in the Common Labs program for the 2018-2019 school year agreed to be a part of the current research project. These teachers received training in specific teaching methods that are used in the Common Labs program, and were supported by the school district. These teachers had approximately 30-32 students per participant teacher, for a total of 180-192 students, out of which approximately 19-21 are students with special needs and approximately 49-53 are ELL students, for a total of 68-74 students in both aforementioned categories. Six 5th grade teachers and their students in science classes who were not participating in the Common Labs program and were engaging in business-as-usual (BAU) science instruction served as control. The BAU classes included approximately 30-32 students per participant teacher, for a total of 180-192 students with approximately the same number and percentage of students who are either ELL and/or received special education services as in the treatment group.

Criteria for Participation

Eligible participants in the present study are science teachers of fifth grade students who were receiving Special Education services and/or were English Language Learners (ELL) within the school district, along with class peers who were Native English speakers (NES) without disabilities. The only selection criterion was: teachers who had in their classroom students participating in either English Language Development (ELD) classes or receiving Special Education services, or both. The research was focused on the population of ELL and students with special needs of this specific school district, and how their science knowledge and academic vocabulary were progressing after being taught the science module either using the Common Labs (CL) or through non-CL methods. Therefore, teachers who had a high number of ELLs and/or students with special needs in their fifth grade science class were recruited.

Matching of Peers

The science classes where the CL program was implemented were matched with classes having similar student demographics from schools that were not participating in the CL program as explained below. During meetings with hierarchically-high administrators of the school district who assigned students to classrooms each year, they used student demographics data, including student SES data on file, to match classes for this study. The district administrators matched peers with similar levels of SES status (percent of the students' participation in the free-and-reduced-lunch program), and student categorization (ELL, students receiving Special Education services for mild/moderate disabilities) from schools that were using the CL program, with similar peers from schools that were not using the CL program. Due to confidentiality concerns, the administrators did not provide these specific data at the student level to the researcher; however, the researcher was present during the peer matching meetings and vetted the criteria the district administrators used. Mean school academic performance level was also used in this matching process. The following table shows the pretest means for each participant classroom in both conditions, to demonstrate the similarity in academic performance at the time of the beginning of the science module.

Table 2
Pretest Means by Classroom and Condition (ordered by students Study ID)

Condition	**CL**	**BAU**
Classroom Means	8.09677	8.32000
	7.06667	10.92308
	7.29630	8.85185
	7.20000	7.34615
	8.36842	6.33334
	9.00640	8.14031
General Mean	***7.83909***	***8.31912***

Power Analysis

To estimate the size of the sample to be included in this study, a statistical power analysis was performed. Using G*Power version 3.1 for ANCOVA (Faul, Erdfelder, Lang, & Buchner, 2007), an *a priori* analysis was conducted with the following parameters: alpha = .05 and power 0.80. The effect size in this study can be estimated as being medium (d = .4) using Cohen's (1988) criteria, which is similar to the effects found in other studies (Cuevas et al., 2005; Li-Grining et al., 2010; O'Connor et al., 2007). The projected sample size needed with this effect size is approximately N=65 students for between group comparison. Thus, the initial sample size of approximately 68-74 students who are ELL and/or receive special education services in the participant science teachers' classrooms was expected to be adequate for the main objective of this study and should have allowed for anticipated ordinary attrition. To ensure the sample size was adequate as per the statistical power analysis performed, the number of teachers invited to participated in either the treatment or the BAU group was greater than 6, to count for the possible rejections. To achieve sufficient power to find effects if they exist, there was the need to include students who were receiving Special Education services together with those who were ELL. It was planned to report the means and SDs of each group (i.e., Special Education, ELL, and Special Education + ELL) separately and combined to determine whether combining data is sensible.

The Common Labs and Comparison Conditions

Several of the aforementioned schools were already using an integrated science program named The Common Laboratories. Although the NGSS were adopted by the California State Board of Education at the end of 2013 (California Department of Education, 2017), California school districts did not start implementing them until a few years later. The participating school district was an early adopter of NGSS, therefore the need of a curriculum that was aligned with the new NGSS was acute. The Common Labs seemed to have helped increase the students' scores on state tests for the first year since its debut in the 2016-2017 school year (as per the school district). After the first year, the administrators of the school district expressed their desire to know specifics of whether students with special needs (who receive special education services and have mild/moderate disabilities) and ELL who are participating in The Common Labs program also improved their academic science vocabulary and knowledge.

The Common Labs program was explicitly created to integrate science with English Language Art (ELA) and mathematics and to use district-endorsed strategies for ELL, as they represented a large portion of the student body in this district – over 25% (see Table 1). Among the teaching strategies employed by the Common Labs science teachers, student collaboration was an important one and constituted the foundation of a well-structured Project-Based Learning (P-BL) setting that was expected to be highly effective for the students. During the instruction time, students needed to talk to members of their groups in order to develop and execute a plan to accomplish the required task. Once this plan was completed, during the final portion of the lesson students needed to write about their experiences during the process. Then, they used mathematics to analyze the data collected and to develop conclusions that have to be justified by evidence gathered during the experimental part of the lesson, which integrated science with mathematics. Moreover, through this integration, students also had some context to help make the learning more meaningful. Particularly for the 5th grade science module this study focuses on, growing radishes hydroponically was the context; NASA was investigating this specific topic and the Common Labs curriculum for 5th grade was addressing this matter as a virtual collaboration of the students with the national agency.

Through the use of innovative curricular practices, the Common Labs program attempts to address different needs that students may have in a typical classroom, whether those are learning English as their second language or having other categories of special needs that can be addressed by teachers during their instruction. The Common Labs approaches students' various learning needs through a wide range of instructional approaches that were shown to be successful in classrooms by a rich body of prior research: P-BL (Bell, 2010; Erdogan & Campbell, 2008; Filippatou & Kaldi, 2010), group work (Baker et al., 2014; Lawrence et al., 2015; Lesaux et al., 2014; Silverman et al., 2013), and scaffolding (August et al., 2014; Hmelo-Silver, et al., 2007; Meyer & Crawford, 2015; Simons & Klein, 2007).

The Common Labs program was first taught at Grade 5 level during the academic year 2016-2017. Teaching the Common Labs was required by the school district but not mandated and the district allowed some latitude in the timeline for implementation. Therefore, the instruction of this module varied from school to school and even among the classrooms within the same school. Thus, the treatment group included 6 fifth grade science teachers who were fully implementing the Common Labs curriculum (and their ELL and/or students who were receiving special education services), while the BAU group comprised fifth grade science teachers who were not using the Common Labs curriculum (and their ELL and/or students who were receiving special education services).

To determine how each curriculum impacted student learning, teachers administered a measure of academic vocabulary in science classrooms, which was the MC vocabulary test that I developed for this study to use as a pretest and posttest. To compare the effects of the Common Labs with the matched control group on science knowledge, it was planned to collect both results from the vocabulary and the benchmark assessment that was district-mandated and supposed to be administered to the entire student population. However, the schools closed in March 2020 due to the pandemic crisis, and did not administer their district-mandated benchmark assessment; therefore, this measure had to be dropped from the analysis, and it will not be mentioned further. In order to understand the gains made by students in the participating teachers' classrooms, the progress of the students in the CL teachers' classrooms was compared with the progress students in the BAU teachers' classrooms had on learning science vocabulary (pre and post administrations of the MC vocabulary test).

Design and Procedure

Students of participating treatment teachers received 8-to-10 weeks of Common Labs instruction, designed for hands-on, integrated science following the fifth grade NGSS and curriculum. Students whose science teachers did not use the CL also received 8-to-10 weeks of science instruction that followed more freely the same state standards as their peers; however, their teachers used more traditional curricula. Two time points were used for the vocabulary test: a pretest at the beginning of the instructional module (beginning of Fall of 2019-2020 school year), and a posttest at the end of the instructional module (end of Spring 2020).

Due to the rapid developments of the COVID-19 pandemic crisis that lead to a complete shutdown of all schools, this research was affected unexpectedly because several posttests were scheduled to be given to the rest of the students after the Spring break of 2020. Once the entire district switched to online teaching/learning, the school district had also a hard time getting the students to take any other tests, quizzes, and assessments, so it decided to drop the district-mandated final unit assessment for the science module. Consequently, this study had a very high attrition rate for which there was the need to compensate using imputation methods for the missing posttest results.

Measures

Multiple-choice vocabulary test. The measure that was used to evaluate students' progress on science vocabulary knowledge was a researcher-designed, multiple-choice (MC) vocabulary test. The 15 items that were used in this assessment are presented in the appendices section (Appendix A).

The multiple-choice science vocabulary test employs words encountered in the NGSS that all classes were required to follow, including the Common Labs curriculum that was created to adhere closely to the NGSS requirements. The vocabulary words for the test were carefully chosen by me in conjunction with the school district administration persons who have decision authority over the science instruction in the district. The vocabulary test was designed to be composed of science vocabulary words that are common to all 5th grade classrooms in the district (whether they participate in CL or not), which are specific to the science vocabulary aligned with the NGSS standards.

The MC vocabulary test was created based on the vocabulary words used by science teachers when teaching the science module about in the *Growing Radishes Hydroponically* that was part of the 5th grade curriculum and also in the science standard on how plants get the materials they need for growth, required in all 5th grade science classes in accordance with NGSS (standard 5-LS1-1). I chose the words using the 5th grade science curriculum in congruence with Biemiller's *Words Worth Teaching* (2010) that emphasizes the importance of building a strong academic vocabulary during elementary grades in order to overcome linguistic barriers that both ELL and students with special needs encounter (Biemiller, 2015). This vocabulary source is used widely for word selection in studies of academic vocabulary. The intention was to measure the adequacy of the Common Labs program at the fifth grade level for improving science learning and science vocabulary of ELL and/or students with special needs. The science vocabulary on this measure sampled vocabulary words introduced to the entire fifth grade student population in the district.

I began by selecting an extended list of words from the district's benchmark unit assessment test. From this list, I eliminated words that were not included in Biemiller's (2010) *Words Worth Teaching*, mainly due to their generally low frequency use. The words chosen are academic words that were to be taught by the teachers of science during the hydroponics module, and are also in the district's unit assessment packet. To test students' knowledge of the taught academic words, the vocabulary test consisted of 15 academic words with 4 answer choices for each word item. From these choices, one is the correct meaning of the word and the other three are various distractors. One distractor contains the meaning of a homonym of the word or parts of it, another contains the meaning of homophones or homographs of the word item, and the third choice is the meaning of a completely different word, primarily the antonym of the word item to be tested. The order of the choices for each item within the test is randomized. Knowledge of these words may benefit the students across their curriculum subject matters, following the NGSS requirement to integrate science subject matter throughout their grade level curriculum (California Department of Education, 2018, pp. 94-129).

To ensure the scores have sufficient range, this measure was administered in December 2018 as a trial pretest in three random science classrooms at one of the elementary schools that was not part of this study. These 5th grade science classrooms consisted of 84 students, out of which 7 were ELL (8.33%), 9 were students with special needs (10.71%), and 68 were typical peers (80.96%). The reason for this pilot test was to find possible ceiling effects for some of the words, and to subsequently replace those words with new ones, that have a higher probability of not being previously known by the students. From the 15 initial items, six (40%) had a percentage of 50 or higher correct responses, possibly due to students' exposure to those words beforehand; therefore, those items were replaced. Similarly to the initial items, the new items were also chosen using Biemiller's *Words Worth Teaching* (2010) and the science curriculum materials that are used district-wide, from the extended list of words created at the beginning of the preparations for this study. The results of the trial pretest can be examined in the table below:

Table 3
Results of the Trial Pretest in 3 Science Classes (Dec. 2018)

	0-25% of students knew the correct definition	26-50% of students knew the correct definition	51-75% of students knew the correct definition	76-100% of students knew the correct definition
Number of words	0	9	4	2
Percentage from total words	0	60%	26.66%	13.34%

The elementary school that agreed to allow the fifth grade science classes to participate in this pilot test is located in a more affluent area of the school district, in which 15.3% students are ELL, compared with 27.7% of students in the same category across the school district (California School Dashboard, 2018). The percentage of students who are ELL is lower than the district average presented previously in Table 1 and the possibility of students being previously exposed to several of the words within the vocabulary test is higher. These were the main reasons for this school not being included in this current study when the criteria for the selection of the participants were discussed with the administrators of the school district; hence, I could include this school in the pilot study.

The 15 items of the MC Vocabulary test are attached as Appendix A. The vocabulary post-test was supposed to be administered electronically by the school district in the spring, but once the schools closed and all students had to participate in online learning, the teachers had a very difficult time making their students take any tests, either mandated by the school district or voluntary (as the posttest of the present study). As a result, there was a very high number of missing posttest results (compared to the pretest), and the district-mandated final unit assessment was dropped by the district completely.

This was the reason for which I have also dropped the district-mandated final unit assessment from this study.

Observation tool. To determine how science was taught in 5th grade classrooms in each condition, the degree to which P-BL was used, and how students who are ELL and/or with disabilities are supported, I used a Classroom Observation Tool that I developed specifically for the present study (Appendix B). This Classroom Observation Tool was used when observing in the classrooms during the 8-10 weeks of the science module that makes the subject of this study. The classroom observation tool includes a rubric that consists of all the elements that the NGSS indicate need to be addressed in science instruction, including: the alignment of lesson activities, bringing out students' understanding of science concepts taught (also named scaffolding methods in the literature), intellectual engagement, use of evidence, Project-Based Learning (application of science), science and engineering practices (SEP, an important aspect of the new state standards) and integration with mathematics, formative assessment, and group work. In addition to the NGSS features, for the purposes of this study I added the following features of classroom instruction and climate: enthusiasm observed separately for students with special needs and for ELL, and the rapport between teacher and students.

Following the detailed rubric, there is a Science Classroom Observation Worksheet that translated the rubric features into scores that were assigned to the classroom activities and teaching strategies used. In parallel with the classroom observation in the treatment classes, there were also conducted observations in the BAU classrooms matched with the treatment classrooms. The teaching strategies and curricular features of the science instruction in the BAU classes that teach the module about growing radishes hydroponically are described using the same Classroom Observation Tool as the science teachers in the Common Labs program.

To complete the Classroom Observation Tool and to establish its reliability, I also planned to video record the classes of the participant science teachers. For this task, an electronic tablet and a smart phone were used for both video and audio recording (depending on how reluctant teachers were in seeing a tablet pointed to their students) that allowed them to be placed in a fixed spot in the classroom while I was observing and taking notes on the Science Classroom Observation Tool sheet. To make sure that I would have the full perception of the classroom activities when using the video or audio recording (for activities that I could have missed during the classroom observation), the teachers provided me with a seating chart of their science classroom, where they made notes of the specific positioning of the students with special needs and those who are ELL while keeping their anonymity. As many teachers (approx. 85%) did not agree to have their students video recorded, the activities in the classroom were audio recorded, while also taking notes about them.

To ensure accurate observation data, interobserver agreement was calculated. One of the observers was a peer of mine from the doctoral program in School Psychology and I was the second observer. I trained my peer on how to use the rubric I developed and how to take notes in the most efficient way while observing the classroom activities. Three lessons per participant teacher were video or audio recorded for the observer/s and the interobserver agreement was calculated on 33% (one third) of the sessions. The ranking of categories on the Classroom Observation Tool was compared between observers and the minimum acceptable/satisfactory level for the interobserver agreement I chose to be 0.8. As Krippendorf (1980) established, a rigorous benchmark for reliability of 0.8 is generally accepted as being a high reliability threshold although many times a lower threshold of 0.75 is also accepted.

The information gathered through the Observation Tool is used descriptively to show whether and how the curriculum differs between CL and BAU conditions. The table below shows the inter-rater reliability (IRR) during the classroom observations in three of the participant classrooms. The scores are for a maximum of 13 choices or features in the rubric.

Table 4
Inter-rater reliability (IRR) scores

Classroom code	No	Yes	IRR
F-035	2	11	0.8461
F-008	1	12	0.923
F-008	3	10	0.7692
F-036	1	12	0.923
F-036	5	8	0.6153
Average IRR			*0.8153*

Research Questions and Data Analyses

1. Do students with special needs (who receive special education services and have mild/moderate disabilities) and students who are ELL who participate in the Common Labs program improve academic vocabulary and general science knowledge more than the matched group of students who are not exposed to this

curriculum, after an 8-to-10-week period of instruction, as shown by their test results?

2. Is the effect of the exposure to the Common Labs program similar for students with special needs, ELL, and their typical peers, as indicated by the test results?

Data Analysis

Because the sample size of the students who receive Special Education services was considered too small for adequate power, depending on effect sizes, the choice was made to combine the categories of students that needed to be analyzed: the categories that comprised the students who receive special education services and those who are ELL. Students who are ELL have similar needs from the point of view of teacher involvement and use of resources, and both have depressed vocabulary scores in comparison to typical learners. The two categories of students both have some special need: the students with disabilities who are receiving Special Education services, as well as the students who are ELL, who also have a special need: in learning English, which is not their first language, and to which some of them did not have enough (if any) exposure before entering school. During the time they are learning the English language, the ELL students need more help from their teachers than the typical students, as well as more resources to be allocated to help them in the process of learning a new language. From a teacher’s point of view, the ELL students need as much extra help as the students who receive Special Education services. The literature supports these needs the ELL students have by acknowledging the high need for the teachers to be trained specifically for teaching science to students who are ELL, as they are also trained to teach science to students who receive Special Education services (Rutt, Mumba, & Kibler, 2019). The high needs of the ELL students surmount the usual learning needs of their typical peers, which makes the teachers need more resources and training to overcome these challenges (Elfers & Stritikus, 2014). Teachers need to use additional resources to help the ELL students learn a new language, while at the same time teaching them the subject matters for that class, along with their classroom peers (typical or with special needs). The use of extra teaching strategies, time and resources for teaching ELL students makes them similar to the students with special needs, from the point of view of this study. Therefore, the choice was made to combine these two categories with the third category of students who are both ELL and receive Special Education services.

For this study, an analysis of covariance (ANCOVA) was used to analyze the data collected, by estimating the variance accounted for by classroom curricula and student classification. ANCOVA investigates the influence that continuous variables (known as covariates) could have on the dependent variable or outcome (Field and Miles, 2010). ANCOVA is comparing two or more regression lines while controlling for variation in the covariates. In ANCOVA for this study, the dependent variable or outcome is the posttest score, while the pretest score is the covariate. The independent variables are the classroom curricula and the student classification. ANCOVA assesses the differences in the posttest score means between the two conditions after accounting for the pretest scores.

The statistical equation of ANCOVA is a combination of ANOVA and linear regression. The equation of a regression line is $Y = a + bX$, where a is the Y intercept and b is the slope (Field and Miles, 2010). The first null hypothesis in ANCOVA is that there are multiple regression lines that have their b slopes equal, therefore they are parallel to each other. The second null hypothesis to be tested in ANCOVA is that the adjusted means of the groups are the same. The equation is presented as: $Y_{ij} = b_0 + b_1X_1 + b_2X_2 + \mathcal{E}_{ij}$ where: Y_{ij} – is the dependent variable (posttest scores) – the j^{th} observation of the i^{th} treatment; b – the slopes; X – the independent variable; $\mathcal{E}_{ij}$ – is the random error in the j^{th} observation of the i^{th} treatment.

The ANCOVA model equation suggests a linear relationship between the outcome and the covariates. The benefit of using ANCOVA for this study is that it can identify effects that could be small and difficult to identify through a different type of analysis; therefore, ANCOVA has greater statistical power. Furthermore, ANCOVA is able to identify and measure patterns in the relations between the dependent variables and the influence of the covariate on the outcome.

As with other statistical models, when conducting ANCOVA, it is important to assess whether the model assumptions are met by the data. These are as follows: the assumption that there are two or more dependent variables that are continuous; the assumption that there are two or more independent variables that consist of two or more categorical groups (here: curriculum and student categorization); the assumption of the existence of a covariate (the pretest scores in this case); the assumption of independence (the observations need to be statistically independent); the assumption of having an adequate sample size; the assumption of multivariate normality (assume that the dependent variables considered together are normally distributed within each group); homogeneity of covariance (the variances in each group of the dependent variables are approximately equal, as well as the correlation between the dependent variables in all groups).

Significant main effects were expected for the two independent variables (receiving CL curriculum or BAU curriculum, and student classification/categorization). It was also expected that the dependent variable (scores on the posttest MC vocabulary test) would be associated with the two independent variables.

Answering the Research Questions

The dependent or outcome variable is the score obtained on the MC vocabulary test. The results were measured as raw scores, later converted to percentages to uniformly illustrate the variables' effect. The independent variables that were tested were the types of instruction used in the treatment and control groups (CL instruction versus traditional instruction in BAU classes), and students' categorization as ELL, students with special needs, ELL and special needs, or typical peers.

The primary outcome was how the participation in either CL or BAU classes and the categorization of students as either ELL or students who receive Special Education services, was associated with the posttest score on the vocabulary test. These students' results were compared with the results of their typical peers.

The null hypothesis would show no difference or very close similarities between the group means of students of the participant teachers across conditions. Contrary to the null hypothesis, after the 8-to-10 weeks of instruction, it is expected that the results will be statistically different across the treatment group and the control (BAU) group. This means that the students of the participant teachers who fully implemented the Common Labs science instruction could show improved scores on the MC vocabulary test.

To answer the first research question – *Do students with special needs (who receive special education services and have mild/moderate disabilities) and students who are ELL, who participate in the Common Labs program improve academic vocabulary more than the matched group of students who are not exposed to this program, after an 8-to-10-week period of instruction, as shown by their test results?* – the dataset obtained from the school district was coded by student status. For this research question the sub-dataset that includes students with special needs and ELL was used and their test results were compared with the test results of same categories of students in the BAU classes. The outcome variable is the posttest score from the MC vocabulary test administered electronically to the students of the participant teachers. The independent variable is the participation in the CL or BAU program. To measure the progress made as a result of the 8-to-10 weeks of instruction in the science module, a baseline/reference score was collected, in the form of the MC vocabulary pretest – in the beginning of Fall of 2019-2020 school year. Descriptive data for each student category by treatment group is shown separately (see tables 8 and 11); however, for analysis students who receive Special Education services, those who are ELL, and those who are both ELL and receive Special Education services were combined in one category to increase the power of analysis. The number of students receiving exclusively Special Education services was very low in both the CL and BAU groups (N=11 students).

Treatment effects of the CL curriculum were tested for the vocabulary knowledge posttest using an ANCOVA with the pretest score as the covariate. Improvement in students' academic vocabulary was measured by comparing the means of their pre-test scores with the posttest scores.

For the first research question, it was expected that the results would show an improvement in the posttest scores of the MC vocabulary test for the students with special needs and those who are ELL, who also receive the CL instruction for 8-to-10 weeks, compared to their peers with special needs and ELL in the classrooms of teachers who are not using the CL curriculum and teaching methods.

To answer the second research question – *Is the effect of the exposure to the Common Labs program similar for students with special needs, ELL, and their typical peers?* – the use of an ANCOVA allowed for an investigation of the results of the Common Labs program on academic vocabulary of typical peers of the ELL and students with special needs, knowing that the participant teachers are all teaching general education classes within which students who receive Special Education services and students who are ELL are mainstreamed. Tests were performed for interaction effects between the two independent variables, curriculum and student categorization and how their combined effect is impacting the dependent variable which is the posttest score on the MC vocabulary test. As follow-up, the univariate results for the dependent variable together with the interaction effect were examined, to determine whether outcomes are similar across the student category that includes the students with special needs and ELL. As with the first research question, the independent variables were the classification as ELL or student with special needs, compared with general education or typical students in the same classrooms where they all receive science instruction.

For the second research question, the null hypothesis was expected to show no significant difference between the 2 groups of students of the participant teachers: students with special needs and ELL (in the treatment group), and typical learners of participant teachers (in the same group). Similarly with the first research question, the analysis for this question is examining the effect of the CL curriculum features on vocabulary knowledge of students who receive Special Education services or are ELL; however, their results were compared with their typical peers who were also receiving the CL curriculum instruction. Performing an analysis that parallels peer groups compares the means of the test results in order to investigate the role of CL curriculum on vocabulary of both students who are ELL and/or receive Special Education services, and their typical peers. It was expected that the means would be statistically similar across the treatment and control groups for each group of learners. This means that the students of the participant teachers who are fully implementing the Common Labs science instruction could show improved scores on the assessments regardless of student category.

Results

Due to COVID-19 pandemic constraints, data collection in the field suffered some modifications. Hence, there was a high attrition rate compared to the number of pretests that were given to the students. Therefore, to be able to use the post-test data missing from the teachers' students corresponding with their pre-test data collected before the shutdown, imputation methods needed to be utilized. From a multitude of imputation methods, hot-deck imputation for continuous data was chosen for this study, after thorough consideration, which I explain below. Because the analyses of the taken samples find the bootstrap confidence intervals near the expected 1.0 (or 0.95), I was able to assume the validity of my study's outcomes (Field & Miles, 2010) . The imputation process is described below.

Missing Data Due to Attrition

The pretest data collected from the field had 239 pretest responses, while the posttest data had only 104 posttests collected. This difference was due to the school district's closing because of COVID-19 lockdown and the education process continuing virtually. Another important factor that determined the high attrition rate was the different schedule that each teacher chose for his/her science module (with the school district's approval). Hence, there were classrooms whose teachers started the science module earlier and were able to finish it and to give their students the posttests before the lockdown was put in place. There was no pattern observed in the various teachers' scheduling of their science unit; even within the same school, some teachers started the science module earlier than others, as the school district gave them full freedom in choosing the schedule for the science unit. Once the lockdown occurred, the rate of attrition increased suddenly for the posttest I developed for this study. Due to these unexpected circumstances, the data missing are considered to be Missing At Random (MAR), a situation that significantly decreases the bias in imputation. I imputed the posttest data missing for students who did not take the posttest.

Table 5

Attrition rates

Original Sample Number (pretest)	Number of actual posttests	Number of imputed posttests
239	104	135

It is worth noting the difference between the pre- and posttest results numbers because of the significant proportion of missing values. Usually, any imputation method would generate a bias because close to 50% of the observations were missing. In cases such as this one, we are presented with limited options: to exclude the cases that have no correspondent posttest data and work only with the pretests that had posttest correspondents, or find a method that would preserve the power of the larger sample and the variances between the values. For this study, I decided to conduct two analyses: on the *Actual Data* (without any imputed posttest results) and another one on the imputed data. The reason for which I chose to conduct these two analyses was to triangulate the results. This way, I could get a more detailed picture of what occurred during this study in the fifth grade science classrooms. Each of the two methods (without and with imputed data) compensates for weaknesses of the other. For the second analysis, the imputation process is described below.

Imputation of the Missing Posttest Results

After evaluating different methods, I decided to use the hot-deck imputation method. The deterministic hot-deck imputation method matches the recipients to their respective donors, usually using the nearest neighbor type. This type of imputation technique is used as the state-of-the-art imputation method for numerous statistical institutes and agencies worldwide, among which the U.S. Bureau of the Census (in different surveys), the National Center for Education Statistics, and the UK Office for national Statistics, to name just a few (Joenssen & Bankhofer, 2012). Moreover, the data being MAR due to unforeseen circumstances previously explained, lead to my choice of using hot-deck imputation for the missing data (posttest results) in my study. The hot-deck imputation method entails identifying the characteristics of the responding students (or *donors*, Andridge & Little, 2010) and associating their pretest scores to the non-responding values (students who had missing posttest scores, or *recipients*, as Andridge and Little named them).

There were two steps: to pair each category to a pretest value and replace the posttest missing values with values of the complete data having similar characteristics. Using this method would imply that most individuals in the same category with the same pretest scores would have the same or very similar post-test scores. A limitation of this method is the assumption that each person within each category would show the same amount of improvement (Myers, 2011). This assumption is probably invalid since there are individual differences for improvement from one student to another, differences that may be dictated by external factors such as physical and emotional factors, besides the teaching process in itself. Another limitation of this method is that the variance of the original data would be affected. To minimize the effect of the imputation on variances, differences between the pretest and the posttest scores for each student were calculated. Then, using the categories that were already established (0 = typical students/peers, 1 = students who are ELL, 2 = students who receive Special Education services, and 3 = students who are both ELL and receive SpEd services), the mean difference of the scores for each category was calculated. These were then used the same way in calculating the means by treatment group. The means of the differences for each category were:

Table 6

Mean differences between pre and posttest scores by student category and condition

Student Category	CL (treatment)	BAU (control)
0 (typical peers)	0.9872	3.0119
1 (ELL)	3.1125	2.7873
2 (SpEd)	1.7778	4.6667
3 (SpEd+ELL)	3.323	2.25

Then, these means were applied to the missing data. Calculations are described below.

Following the above process, if a student was in category 0 and the CL group, the posttest score for him/her was calculated by adding 0.9872 to the pretest score; for category 0 in BAU group, 3.0119 was added, and so on (see Table 7 above). This means that students in the same category would have similar variations between pretest and post-test scores. Then, the minimum and maximum scores were restricted to 0 and 15, respectively, which are the minimum and the maximum possible scores to the vocabulary pre- and posttests (see Appendix A).

Treatment condition (CL):

To show the differences between the pretest and posttest scores, the table below presents the descriptive statistics for the CL. Means and standard deviations are included.

Table 7

Descriptive statistics for pretest and posttest (before and after imputation) by student category – CL:

Student Category	Score means (pretest)	Score SD (pretest)	Score means (post, actual)	Score SD (post, actual)	Score means (post, imputed)	Score SD (post, imputed)
0	9.05	3.0017	10.03	3.0916	9.76	3.0527
1	6.03	2.4935	9.14	4.5617	9.2	2.666
2	4.75	2.2519	4.66	2.0816	4.07	1.471
3	4.22	1.3944	6.00	2.3094	6.42	1.7072
Average	*6.01*	*2.2853*	*7.45*	*3.011*	*7.36*	*2.2243*

Below, descriptive statistics are presented for the control condition (BAU), for the pretest and the posttest scores before and after imputation, for both actual and imputed samples.

Control condition (BAU):

Table 8

Descriptive statistics for pretest and posttest (before and after imputation) by student category – BAU:

Category	Score means (pretest, actual)	Score SD (pretest, actual)	Score means (post, actual)	Score SD (post, actual)	Score means (post, imputed)	Score SD (post, imputed)
0	9.67	2.2287	12.19	1.6999	12.01	2.0098
1	6.95	1.6382	9.84	2.7339	8.18	3.8924
2	4.33	1.5275	9	1.4142	7.55	2.6978
3	5.41	2.6097	8.42	1.1338	6.17	2.9131
Average	*6.61*	*2.0010*	*9.86*	*1.7454*	*8.47*	*2.8782*

Data Analysis for *Actual Data* Sample

The first data analysis conducted was on the sample called the *Actual Data*. These data consisted of original, measured students' scores; there were no imputed mean scores. The sample size for this data analysis was 104. Below, the summary of this sample is presented:

Table 9
Frequencies for the Actual Data sample by condition

	N	%
CL (treatment)	40	38.5
BAU (control)	64	61.5
Total	**104**	**100.0**

Table 10
Frequencies for the Actual Data sample by category

	N	%
Typical peers	68	65.4
ELLs	20	19.2
SpEd	5	4.8
SpEd+ELL	11	10.6
Total	**104**	**100.0**

The descriptive statistics for the *Actual Data* sample are presented below:

Table 11
Descriptive statistics for the Actual Data pre- and posttest scores

	Pretest scores	Posttest scores
N	104	104
Mean	8.52	10.41
Median	9.00	11.00
Std. deviation	2.946	3.016
Min.	2	2
Max	14	15

Testing for the within-subjects effect, the results show the p-value $p=.047$ when analyzing for test by treatment, which shows a significant effect ($p < .05$), $F(1,100) =$ 4.036 and it accounts for 29.9% of the variance (*partial* $Eta^2 = 0.299$). These results show a significant, positive correlation between the science module teaching and the posttest results. In this analysis, the group of SpEd+ELL students in the treatment condition (CL) is compared with the same group in the control condition (BAU) and the results show that the control group had a greater improvement than the treatment group. Similarly, the analysis shows an F-statistic $F(1,100) = 4.711$ and a p-value of 0.032 for test by category which translates into a positive correlation between the science module teaching and the posttest scores, and a significant effect ($p < .05$) of the test on each of the categories. In this analysis, the categories analyzed are SpEd+ELL against typical peers in both conditions. To reiterate, the test by treatment by category 3-way interaction was not significant ($p = .195$). Thus, when the four groups (SpEd+ELL treatment, SpEd+ELL control, typical treatment, typical control) were compared, the differences in growth were not significant for the 3-way interaction. In contrast, the test by category interaction was significant ($p<.05$). This means when the growth for SpEd+ELL is compared against growth for typical students, the SpEd+ELL group showed a higher growth than the typical peers group, regardless of condition. The test by treatment interaction was also significant ($p<.05$), which shows that the growth for the control group was significantly greater than the growth for the treatment group.

Table 12

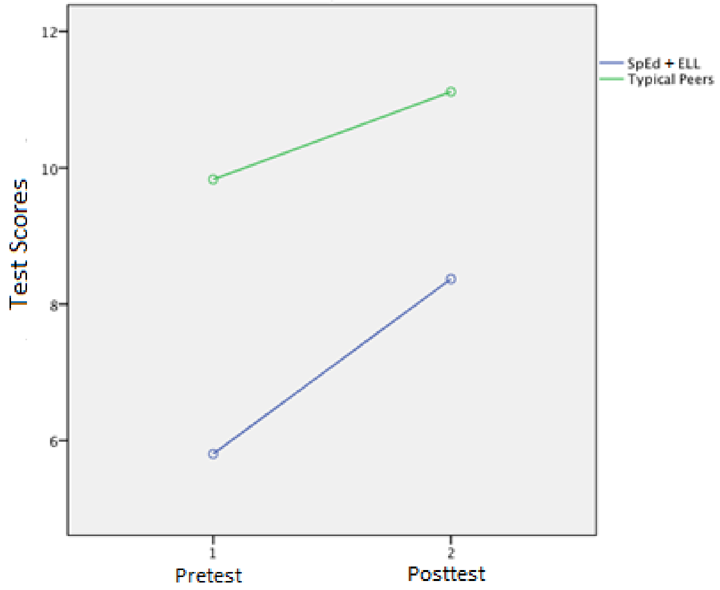
Test of within-subjects effects for science module on posttest results of combined student categories – Actual Data

	Sum of Squares	df	Mean Square	F	p-value
Test by treatment	15.673	1	15.673	4.036	.047
Test by category	18.294	1	18.294	4.711	.032
Test by treatment by category	6.607	1	6.607	1.701	.195

Next, a line graph shows combined typical peers in both treatment (CL) and control (BAU) against the combined treatment and control groups for combined SpEd, ELL, and SpEd+ELL. Analyzing combined Sped+ELL in both CL and BAU, it can be noticed a greater improvement for Sped+ELL (slope is steeper) than their typical peers. This analysis is not separating the two conditions; it shows the performance of the SpEd+ELL group versus the typical peers group, in both conditions. Regardless of which condition students were in, the students who were ELL and received Special Education services had bigger gains than the typical peers group. The starting points of the two slopes show a big difference, favoring typical peers, which was expected when comparing typical peers with SpEd+ELL before starting the science module due to multiple factors that will be discussed in the next section.

Figure 1

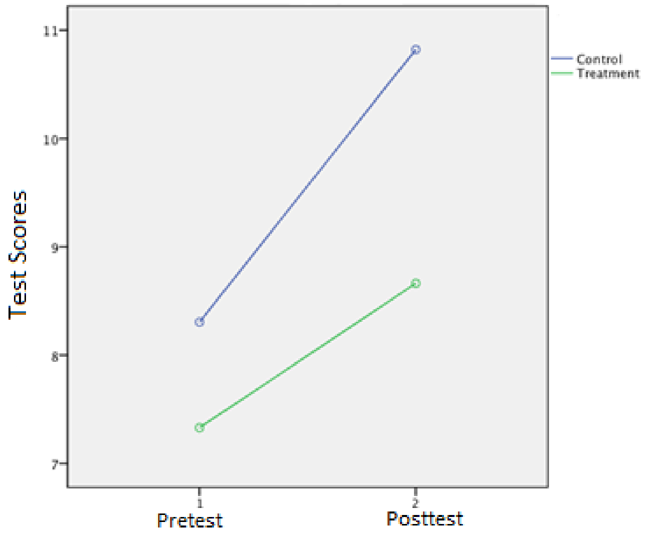
Difference between pre- and posttest scores for SpEd+ELL vs typical peers – Actual Data



Following, a line graph shows comparison between the conditions (CL against BAU). This analysis does not separate the students by category; it shows the performance for the treatment group as a total against the control group as a total as well. The starting points for both conditions are less than a point away so the difference is not as great as in the previous comparison, considering the 15 possible points total. This line graph shows the BAU group having a much bigger improvement than the CL group. Since the joint effect of test treatment by category was not significant (Table 17), the 3-way interaction results are not included here.

Figure 2

Difference between pre- and posttest scores for treatment (CL) vs. control (BAU) – Actual Data



Data Analysis for *Imputed Data* Sample

As previously mentioned on page 44, I decided to conduct a second analysis on the *Imputed Data* sample, and this is presented next. The second analysis conducted used the sample that contains the imputed posttest results, named the *Imputed Data*. The *Imputed Data* sample was analyzed to offset for a possible type 2 error, which could occur in the first analysis due to the small sample of *Actual Data* (e.g., there could be a significant effect that could go unnoticed due to the small sample size). To start with the data analyses on the sample of 239 students of the participant teachers, below is the classification by group and student category. The sample size was 239, which included the *Actual Data* and the *Imputed* posttests for the missing ones. The analysis is presented below:

Table 13

Frequencies for the Imputed Data sample by condition, including students for which posttest results were imputed

Condition	N	%
CL (treatment)	125	52.3
BAU (control)	114	47.7
Total	**239**	**100.0**

Table 14

Frequencies for the Imputed Data sample by category, including students for which posttest results were imputed

Category	N	%
ELL	50	20.9
SpEd	11	4.6
Sped+ELL	21	8.8
Typical	157	65.7
Total	**239**	**100.0**

The descriptive statistics for the *Imputed Data* sample are as presented below:

Table 15

Descriptive statistics for the Imputed Data pre- and posttest scores

	Pretest scores	Posttest scores
N	239	239
Mean	8.13	10.10
Median	8.00	10.00
Std. deviation	3.063	3.087
Min.	2	2
Max	15	15

After combining the 3 categories of students: students in the Special Education category, students who are ELL, and students who are both ELL and receive Special Education services, the descriptive statistics for the *Imputed Data* are shown below:

Table 16

Descriptive statistics for Imputed Data per condition and category

	Category revised	Condition	Mean	SD	N
Pretest scores	SpEd+ELL	BAU	6.19	2.221	32
		CL	5.50	2.384	50
		Total	***5.77***	***2.332***	***82***
	Typical peers	BAU	9.77	2.279	82
		CL	8.92	2.958	75
		Total	***9.36***	***2.651***	***157***
Posttest scores	SpEd+ELL	BAU	9.09	2.115	32
		CL	7.86	3.017	50
		Total	***8.34***	***2.754***	***82***
	Typical peers	BAU	12.27	1.994	82
		CL	9.65	3.025	75
		Total	***11.02***	***2.850***	***157***

The analysis was conducted for the full model, to test if the change in posttest scores from the pretest scores for typical peers differed from the change for students who are ELL and/or receive Special Education services, and if they differ also by condition (CL vs. BAU). So, as a result, the 3-way interaction tests whether the treatment had a different effect for typical students versus students in the combined category of ELL and Special Education. The test for the 3-way interaction was needed to show the big picture of what occurred during the study.

When testing for within-subjects effects, the results were as follows:

Table 17

Test of within-subjects effects for science module on posttest results of combined student categories – Imputed Data

	df	Mean Square	F	p-value
Test by Treatment	1	34.838	19.052	.000
Test by Category	1	26.913	14.719	.000
Test by Treatment by Category	1	9.699	5.305	.022

The test by treatment by category joint effect is significant in the case of *Imputed Data* (*F(1,235)* = 5.305; $p < 0.05$), meaning the treatment had a stronger effect for students who are ELL and/or receive Special Education services, compared to their typical peers. In addition, the treatment and category of students separately each had a significant effect on their improvement from pre to post test ($p < 0.001$). The joint interaction of test by treatment by group was significant for all groups; among these groups, the typical peers in the CL group showed the least improvement in posttest scores.

To verify these findings, a power test of between-subjects effect was conducted, and the results are shown below:

Table 18

Test of between-subjects effects – Imputed Data

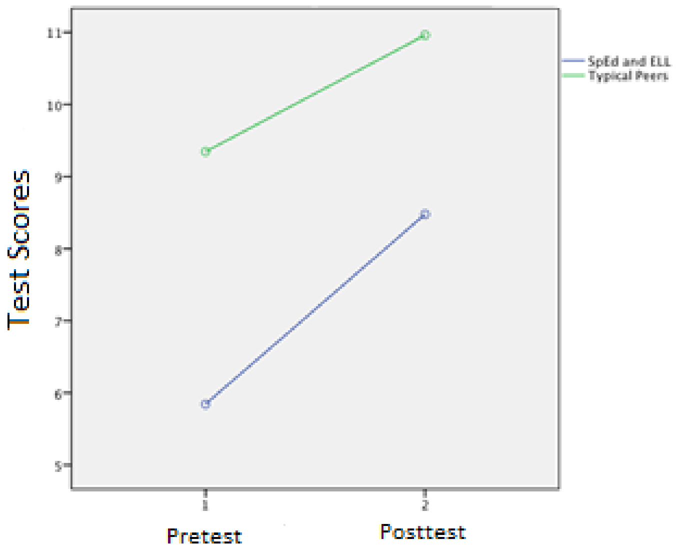
Source	Observed Power
Test by Treatment	.983
Test by Category	1.000
Test by Treatment by Category	.215

Compared to a standard power of .8 (or 80%), the results above show that both *test by treatment* and *test by category* have strong power, while test by treatment by category (the 3-way joint interaction) has small power and, thus, a small chance of having a significant correlation with the posttest results (.215 or 21.5%).

Following, a line graph shows the progress of the two categories of students, the combined SpEd+ELL against their typical peers for the *Imputed Data*. The difference between the initial scores is more than 3 points, which was expected, as the special needs group performed lower in pretest. This analysis did not separate the two conditions, but it shows the performance of the SpEd+ELL group as a whole versus their typical peers group in both conditions combined. It can be noticed by the slope of the two groups that the students who were ELL and received Special Education services had higher gains than the typical peers group, regardless in which condition they were placed.

Figure 3

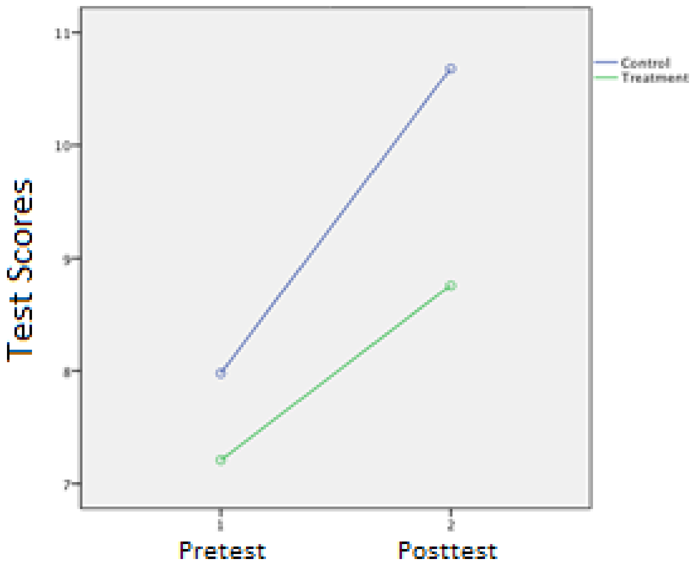
Difference between pre- and posttest scores for SpEd+ELL vs. typical peers by condition – Imputed Data



The next line graph shows the progress of the students in both groups by condition, also within the *Imputed Data* sample. This analysis does not separate the students by category; it rather shows the performance for the CL group as a total against the BAU group as a total as well. The starting point for both conditions is less than a point away so the difference is not as great as in the previous comparison considering 15 possible points. This line graph shows the BAU (control) group having a much higher improvement than the CL (treatment) group.

Figure 4

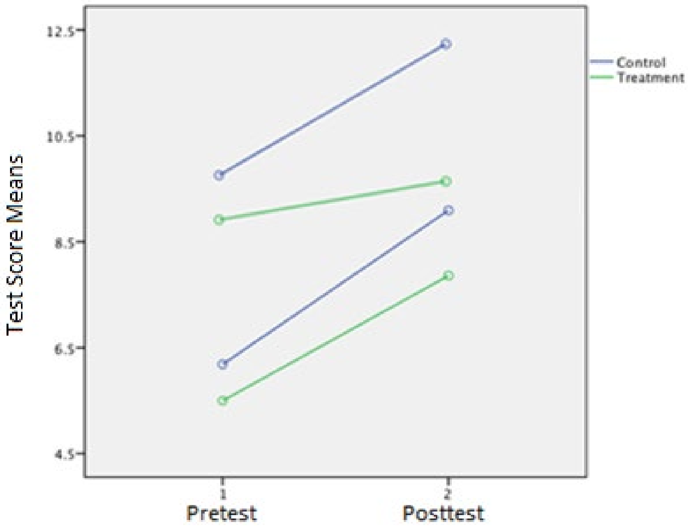
Difference between pre- and posttest scores for CL vs. control BAU – Imputed Data



After analyzing the two samples, some patterns can be noticed. These will be addressed next. First, the line graphs for the change in test scores means from pre- to posttest within the *Actual Data* sample are presented:

Figure 5

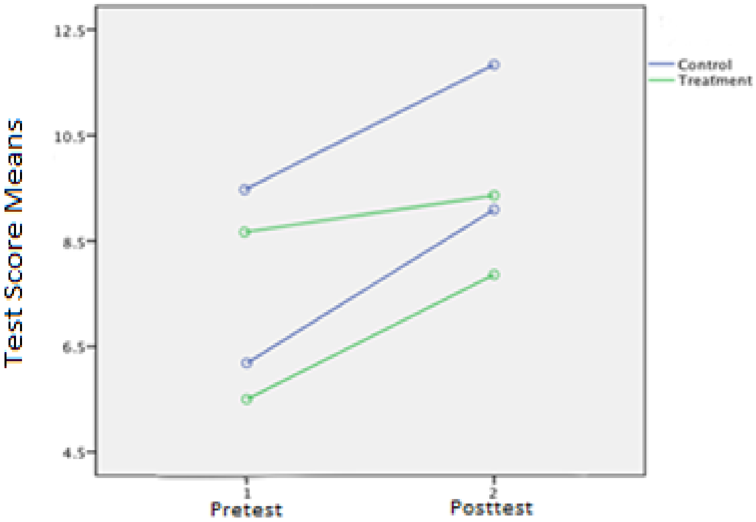
Difference between pre- and posttest scores means for typical students (upper) and SpEd+ELL (lower) by condition – Actual Data



Next, the line graphs for the difference/change in test scores means from pre- to posttest within the *Imputed Data* sample are presented:

Figure 6

Difference between pre- and posttest scores means for typical students (upper) and SpEd+ELL (lower) by condition – Imputed Data



As it can be noticed, the Y axis is similar for all the graphs, depicting the test scores means, which makes it easier to compare the test scores for both samples and all categories analyzed.

- SpEd+ELL – treatment (CL)
- SpEd+ELL – control (BAU)
- Typical students – treatment (CL)
- Typical students – control (BAU)

The patterns detected in the difference/change of the scores from pretest to posttest for these groups are as follows:

- The pattern of change within the sample of *Actual Data* is almost identical with the one within the *Imputed Data* sample. This was expected, knowing that almost half of the students are the same in both samples.
- The pattern of change for the SpEd+ELL students (in both *Actual Data* and *Imputed Data* samples) for CL and BAU groups is very similar. This suggests that the treatment for SpEd and ELL had no significant effect when compared with the BAU; however, there was a bigger improvement in test scores for students in the Sped+ELL group as a whole, regardless of whether they were in CL or BAU.
- The pattern of change for typical peers (in both *Actual Data* and *Imputed Data* samples) showed a greater gain for typical students in BAU as compared to CL.

Overall, there is a noticeable pattern for both samples, *Actual Data* and *Imputed Data*: all groups show similar gain from pretest to posttest scores, with the exception of the typical peers in CL. These students showed less gain than each of the other 3 groups.

Classroom Observation Tool

Information from the Observation Tool was summarized by condition (CL or BAU) and tabled to consider differences across curricula. When the classroom observations were conducted, 4 classrooms in the CL group and 3 in the BAU group were observed before the Spring break. The rest of the classroom observation sessions were scheduled to occur after the Spring break but schools closed at that time due to the pandemic regulations. For interrater reliability, there were 5 observations conducted in 3 classrooms by 2 observers, in parallel (as described in the *Observation Tool* sub-section above). In total, 18 classroom observations in CL classes and 10 observations in BAU classes were conducted. More were scheduled to occur after the spring break, but that was when the pandemic lockdown started, and all schools switched to online teaching, so no other classroom observations were conducted. Below, the 13 features of the *Classroom Observation Tool* features are presented, to illustrate potential differences during instruction in the two conditions.

Table 19

Classroom Observation Tool mean scores by condition (score means from 0 to 6)

Rubric feature	CL condition	BAU condition
1. Alignment of lesson activities	5.38	5.25
2. Scaffolding (bringing out science concepts)	5.45	4.75
3. Intellectual engagement	5.2	5.0
4. Use of evidence	5.2	4.75
5. P-BL methods (Application of science)	5.33	4.8
6. SEP and integration w. other subjects	4.01	2.5
7. Formative assessment	4.84	4.83
8. Group work	5.54	5.4
9. Teacher scaffolding - students in SpEd	4.33	4.67
10. Teacher scaffolding - students ELLs	5.2	5.75
11. Enthusiasm - students in SpEd	3.87	4.0
12. Enthusiasm - students ELLs	5.07	5.16
13. Rapport between teacher & students	5.24	5.83

The findings captured in this tool will be addressed in the *Discussion* section.

Besides the thirteen features observed and scored in the rubric, there were also observations of the integration of science with mathematics and English Language Arts (ELA), activities involving technology, group activities, individual work, and whole class instruction. These activities were noted in the rubric as number of occurrences and the amount of time they were used. The raters noted that the integration of science with other subjects occurred in 77% of the classrooms in the CL group and 61% in the classrooms in the BAU group. For the rest of the activities, the differences between the two conditions were similar.

Teacher Effects

To evaluate teacher effects on students' learning, a teacher survey was developed, in which teachers are asked to give information about their level of education, length of teaching within the school district, length of teaching prior to the implementation of the CL in the district, and a few questions about their own teaching methods. Data from the surveys was supposed to be summarized in a table by curricular condition to explore whether differences in teacher characteristics across conditions should be considered in the analysis. Due to switching to remote learning, the district provided the information requested in the survey due to the impact the pandemic had on teachers' time with preparation for online schooling.

Table 20

Results of the District Information on Teachers' Experience and Skills

Teacher Survey	CL	BAU
Level of education	BA, MA	BA, MA
Total number of years being employed as a teacher	>5 years	>5 years
Number of years as a teacher before CL was implemented	>3 years	>3 years

Discussion

The main objective of this study was to examine the effects of the new P-BL science program, called Common Labs, or CL, that followed closely the NGSS standards on fifth graders' academic vocabulary improvement. In particular, this study examined the gains students made between the pretest and posttest vocabulary scores first between groups of students who were ELL and/or received Special Education services (SpEd+ELL) in CL and the control (BAU) group. In a second analysis, the gain in academic vocabulary of students who were SpEd+ELL from the CL group was compared against the gain of their typical peers in the same classes.

Numerous studies described the importance of addressing academic vocabulary in teaching science, especially for students who are ELL and/or who receive Special Education services (Bravo & Cervetti, 2014). Studies have also suggested advantages for integrating science with other subjects using teaching strategies that intertwine these subject matters into one that could improve the academic outcomes for students who are ELL and with special needs, as well as for their typical peers (Meyer & Crawford, 2015;

Pine et al., 2006). Furthermore, the literature has suggested that scientific inquiry introduced as Project-Based Learning (P-BL) could and is being used to advance science academic vocabulary of students (August, Branum-Martin, Cardenas-Hagan, & Francis, 2009; Bravo, Cervetti, Hiebert, & Pearson, 2007).

The P-BL science unit scrutinized in this study was designed to improve the science academic vocabulary of students who are ELL and/or receive Special Education services as well as their typical peers by using all the aforementioned teaching practices. The main findings of this study can be summarized as follows: (1) students who were ELL together with those who received Special education services in BAU classes had higher improvement in their academic science vocabulary than in CL classes, and (2) in the CL, academic vocabulary gains were higher for students who received Special Education services and/or were ELL than for their typical peers.

Due to the pandemic situation, in which schools closed and the process of education continued only virtually, attrition was large. As a result, I conducted analyses with the *Actual* sample, and also imputed missing data to estimate results with a larger sample. The findings were consistent within both the *Actual* (smaller) sample and *Imputed* (larger) sample. Regardless, students in the combined SpEd+ELL groups in both conditions grew more in academic vocabulary than their typical peers in the same classes. That the *Actual* and the *Imputed Data* show such similarity strengthens the use of imputation in this case, suggesting that if the *Actual Data* sample had been larger, it might have shown significance in the 3-way interaction. Thus, it is possible that the 3-way interaction was not significant in the *Actual Data* sample due to a type 2 error. The fact that the means for the two samples (*Actual* and *Imputed*) are so similar also supports the idea that the imputation used provided good estimates of the missing posttest scores. The results showed that the hypothesis that the new CL, P-BL-based program, would have better results for vocabulary than the BAU group was not supported.

I considered whether the nonsignificant effects of the CL could be due to lack of training, or poor implementation of CL, so I collected this information from the district. As per the school district's administrators, their teachers were all trained in specific teaching techniques to be utilized with students with special needs. Additionally, the teachers in this school district were all fully credentialed (no temporary teaching permits) and had a Bachelor in Arts degree at the least, with several teachers holding a Master's in Arts degree. Per their records, the school district also provided several training sessions yearly for all their teachers to keep their knowledge updated regarding the teaching techniques for students with various needs, including students who are ELL, and specific teaching methods to be used in science classrooms. Moreover, teachers using CL received additional training on using P-BL in their classrooms and how to integrate science with other subject matters, especially with mathematics. The only difference in the inservice provided to teachers by the district in the two conditions was the additional training provided to the teachers in the CL group.

Due to the additional training in CL, it was expected to observe strong implementation of CL, higher levels of group work and scaffolding in CL classes, and higher improvement in science vocabulary of the students who received the new CL program. Instead, the data analyses showed that the students in the BAU group performed with higher posttest results (Figure 5). It is possible that the district training for support of high-needs students and for science instruction in general minimized any potential effects of the new science curriculum; however, I did not observe this training, so can only speculate.

Classroom Observation Tool

The Classroom Observation Tool collected compelling evidence to help explain the findings. Using this descriptive tool, the observers took notes to document potential impact of classroom activities on student outcomes. Although this tool documented how teachers interacted with SpEd+ELL, it was not designed to capture differences between support for these students compared to their typical peers. Nevertheless, both observers noted that teachers in the CL group spent a lot of their time with SpEd+ELL students. This observation, paired with the highly demanding nature of conducting P-BL teaching in a fairly new program, could have left the typical students with less help than in traditional classrooms. In contrast, the BAU teachers, who also gave much of their time to SpEd+ELL students, did not insist that their students do all the hands-on work themselves, since teachers performed the experiment as a demonstration in front of the classroom. Considering the means of the scores obtained from classroom observations (Table 18 in *Results* section), one can notice few apparent differences across conditions. This could show that, while teachers in the CL group received more training on how to apply the P-BL techniques in classroom, they fell short in applying these specific methods. The CL program was still new at the time, and perhaps needed more focused training than these teachers received. Teachers in both conditions appeared to be using specific practices from their training to scaffold understanding of students who were SpEd+ELL. While group work could improve students' analytical and problem-solving skills, as Harskamp and Suhre (2006; 2007) have found, analytical skills were not documented on the Observation Tool nor tested via vocabulary scores. It is possible that overall science understanding was not captured by the vocabulary test, and the district science test was not administered due to the pandemic. Thus, it is possible that outcome differences went undetected by the measure used here.

It was surprising that the teacher enthusiasm score for students receiving Special Education services was among the lowest score means. In contrast, the whole rapport between teachers and students received high scores, showing the great involvement of teachers across conditions in their classroom interaction with students, which would support the specific training that all teachers received to improve their skills in working with students who were SpEd+ELL.

One possible difference across conditions was integrating subject matters, a feature other studies have suggested improves science outcomes while also helping students to develop a solid academic vocabulary (August et al., 2009; Brophy, Klein, Portsmore, & Rogers, 2008; Santau, Maerten-Rivera, & Huggins, 2011; Vadasy, Sanders, & Nelson, 2015). In observing the CL classes, 84.2% of the sessions demonstrated integration of science with mathematics and/or English Language Arts. However, only 66.67% of BAU observations included this feature. The observations of the number of occurrences and time spent in these activities showed fewer integration activities used in the BAU group compared with the CL group. However, the score of these activities in the CL group was still not as high as expected, in view of the additional training these teachers received before and during the school year.

From the observations of all features observed (e.g., integration of science with mathematics and ELA mentioned above, activities involving technology, group activities, individual work, and whole class instruction), including recordings of the number of occurrences and the time they were used, it can be speculated that the teachers in the CL condition were not using the CL methods to their full extent, or were not balancing their utilization as instructed. At the same time, teachers in the BAU group, while not using much science integration with other subjects or student-centered group work during the experiments, had an overall better outcome in vocabulary scores. Conjecturing, this could show that using teaching techniques well-known to them gave teachers in the BAU more time to teach science.

Limitations

As a consequence of the restrictive pandemic crisis circumstances, when all schools shut down for in-person teaching and education moved to online schooling, the attrition this study suffered was considerable. Once the teaching moved to a virtual model, many students did not perform as expected. The Vocabulary posttest for this study suffered high rates of attrition because teachers could not make taking the vocabulary test compulsory in classes that had not already completed it.

A major loss for this study due to the school district shutdown was the lack of a general science measure. The district-mandated science unit final assessment that was supposed to be administered to students at the end of the school year, when all the teachers would have finished their science module, was not administered. This district mandated test would have provided scores to show students' improvement in their science knowledge and analytical thinking skills. These scores, enhanced with the science vocabulary scores (Appendix A), would have depicted a more detailed estimate of how teachers in both conditions performed in their teaching of science to various categories of students, using different types of programs and teaching techniques.

Another limitation of the current study could be regression to the mean for scores of students in the SpEd+ELL group. These students had lower scores than their typical peers at pretest; therefore, they might have had more room to grow from the initial pretest to the posttest scores (Figures 1, 3 and 5), accounting for the higher gains of SpEd+ELL.

A third limitation is the choices for what to document on the Observation Tool. While the Classroom Observation Tool reported the time spent in different classroom activities, there was no record of the total science instructional time. A record of total science time could have shown differences in opportunity for students to learn science vocabulary. Teaching excellence is another feature that was not monitored in the Classroom Observation Tool. It is possible that differences in overall instructional skill teaching science might have influenced outcomes in the current study.

The pandemic also limited classroom observations to 58.3% of the teachers that were scheduled to be observed. According to the schedule, the raters observed 66.7% of the teachers in the CL condition and 50% of the teachers in the BAU condition. Considering this, the means of the scores in the Classroom Observation Tool might have been different, had all the teachers been observed.

Conclusions

The joint effect of test by treatment by category shows significant improvement on the vocabulary posttest scores for all groups in the study (Table 16). This improvement suggests that teaching the science unit had a positive impact on vocabulary growth in both conditions and all categories of students. Nevertheless, no effect was found supporting CL over traditional science instruction. When scrutinizing the effect on the outcome for each of the groups, the least improvement was displayed by the typical students in the CL group (Figures 5 and 6). Although the features of instruction documented on the Observation Tool were similar across groups, teachers in the CL program were implementing a new curriculum. The effort they expended to learn the new method may have negatively affected their typical students.

The school district provided the teachers who were starting the CL with supplementary training over the summer and regular refresher sessions during the school year; however, it may be that teachers in CL classrooms were not using the newly-received training to its full extent. It is possible that their training lacked ways to balance all the P-BL methods with other classroom activities and traditional teaching strategies, leaving this balance to the latitude of each teacher. It was clear during the observations that some of the teachers in the CL group were more successful than others in applying specific CL teaching techniques. At the same time, teachers in the BAU group continued the science teaching to which they were accustomed and experienced. Perhaps this experience contributed to the greater improvement in science vocabulary scores for their typical students.

With experience, differences in instruction could become apparent where it was not documented in the current study. A broader measure of science knowledge than vocabulary might also have found outcome differences. The district had planned to administer such a measure; however, the pandemic interfered with its administration. Nevertheless, it may also be that CL is simply not an improvement over more traditional instruction.

REFERENCES

Altemeier, L., Jones, J., Abbott, R. D., & Berninger, V. W. (2010). Executive functions in becoming writing readers and reading writers: Note taking and report writing in third and fifth graders. *Developmental Neuropsychology, 29*(1), 161-173.

Amos, S., & Boohan, R. (2003). *Aspects of teaching secondary science: perspectives on practice*. Routledge.

Andridge, R. R., & Little, R. J. A. (2010). A review of hot deck imputation for survey non-response. *International statistical review*, *78*(1), 40-64. doi:10.1111/j.1751-5823.2010.00103.x.

Apthorp, H., Randel, B., Cherasaro, T., Clark, T., McKeown, M., & Beck, I. (2012). Effects of a supplemental vocabulary program on word knowledge and passage comprehension. *Journal of Research on Educational Effectiveness, 5*(2), 160-188.

August, D., Branum-Martin, L., Cardenas-Hagan E., & Francis, D. J. (2009). The impact of an instructional intervention on the science and language learning of middle grade English language learners. *Journal of Research on Educational Effectiveness*, *2*, 345–376. DOI: 10.1080/19345740903217623

August, D., Branum-Magnum, L., Cardenas-Hagan, E., Francis, D. J., Powell, J., Moore, S., & Haynes, E. F. (2014). Helping ELLs meet the Common Core State Standards for literacy in science: The impact of an instructional intervention focused on academic language. *Journal of Research on Educational Effectiveness*, 7, 54–82.

Baker, S., Geva, E., Kieffer, M. J., Lesaux, N., Morris, J., Proctor, P., . . . McCallum, D. (2014). *Teaching academic content and literacy to English learners in the elementary and middle school*. Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. Retrieved on May 9, 2017 from NCEE website: http://ies.ed.gov/ncee/wwc/

Beck, I. L., McKeown, M. G., & Omanson, R. C. (1987). The effects and uses of diverse vocabulary instructional techniques. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 147-163). Hillsdale, NJ, US: Lawrence Erlbaum Associates.

Bell, S. (2010). Project-based learning for the 21st century: Skills for the future. *The Clearing House, 83*, 39-43.

Biemiller, A. (2010). *Words worth teaching: Closing the vocabulary gap*. Columbus, OH: McGraw-Hill SRA.

Biemiller, A. (2015). Which Words Are Worth Teaching? *Perspectives on Language and Literacy*, *41*(3), 9.

Bravo, M. A., & Cervetti, G. N. (2014). Attending to language and literacy needs of English Learners in science. *Equity & Excellence in Education, 47*(2), 230-245.

Bravo, M. A., Cervetti, G. N., Hiebert, E. H., & Pearson, P. D. (2007). From passive to active control of science vocabulary. *56th Yearbook of the National Reading Conference*. (pp. 122–135). Chicago, IL: The National Reading Conference.

Bravo, M. A., & Garcia, E. E. (2004). Learning to write like scientists: English language learners' science inquiry and writing understanding in responsive learning contexts. *National Clearinghouse for English Language Acquisition*, *1*, 1–20.

Brewer, C. (2004). Near real-time assessment of student learning and understanding in biology courses. *BioScience, 54*(11), 1034-1039.

Brophy, S., Klein, S., Portsmore, M., & Rogers, C. (2008). Advancing engineering education in P-12 classrooms. *Journal of Engineering Education, 2008*, 369-387.

Bruhn, A. L., Hirsch, S. E., & Lloyd, J. W. (2015). Treatment integrity in school-wide programs: A review of the literature (1993–2012). *The Journal of Primary Prevention*, 36(5), 335-349.

Bruhn, A. L., Lane, K. L., & Hirsch, S. E. (2014). A review of Tier 2 interventions conducted within multitiered models of behavioral prevention. *Journal of Emotional and Behavioral Disorders, 22*(3), 171-189.

California Department of Education (2018). *2016 Science framework for California public schools Kindergarten through grade twelve*. Retrieved from https://www.cde.ca.gov/ci/sc/cf/documents/scifwchapter4.pdf.

California Department of Education (2017). *NGSS for California Public Schools, K-12*. Retrieved from https://www.cde.ca.gov/pd/ca/sc/ngssstandards.asp.

California School Dashboard (2018). Retrieved from https://caschooldashboard.org.

Cohen, E. G. (1990). Teacher as Supervisor of Complex Technology. *Theory into PRactice* , 78-84.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin, 70*(4), 213-220.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. *Second Edition.* New York, NY: Lawrence Erlbaum Associates.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: teaching the craft of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Hillsdale, NJ: Erlbaum.

Cuevas, P., Lee, O., Hart, J., & Deaktor, R. (2005). Improving science inquiry with elementary students of diverse backgrounds. *Journal of Research in Science Teaching, 42*(3), 337-357.

Dalacosta, K., Kamariotaki, M., Palyvos, J. A., & Spyrellis, N. (2009). Multimedia application with animated cartoons for teaching science in elementary education. *Computers & Education, 52*(4), 741-748.

Doppelt, Y. (2005). Assessment of Project-Based Learning in a Mechatronics context. *Journal of Technology Education, 16*(2), 7-24.

Elfers, A.M., & Stritikus, T. (2014). How school and district leaders support classroom teachers' work with English Language Learners. *Educational Administration Quarterly, 50*(2), 305-344.

Erdogan, I., & Campbell, T. (2008). Teacher questioning and interaction patterns in classrooms facilitated with differing levels of constructivist teaching practices. *International Journal of Science Education, 30*(14), 1891-1914.

Ertmer, P. A. & Simons, K. D. (2006). Jumping the PBL implementation hurdle: Supporting the efforts of K-12 teachers. *Interdisciplinary Journal of Problem-Based Learning, 1*(1), 40-54.

Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.

Field, A., & Miles, J. (2010). *Discovering statistics using SAS*. Los Angeles, CA: SAGE.

Filippatou, D., & Kaldi, S. (2010). The effectiveness of project-based learning on pupils with learning difficulties regarding academic performance, group work and motivation. *International Journal of Special Education, 25*(1), 17-26.

Foorman, B. R., & Torgesen, J. (2001). Critical elements of classroom and small-group instruction promote reading success in all children. *Learning Disabilities Research & Practice, 16*(4), 203-212.

Fradd, S. H., Lee, O., Sutman, F. X., & Saxton, M. K. (2002). Promoting science literacy with English Language Learners through instructional materials development: A case study. *Bilingual Research Journal, 25*(4), 479-501.

Ghosh, B., Hokom, M., Hunt, Z., Magdaleno, M., & Su, B. (2008). *Bilingual education in Texas: Exploring best practices. Final Report.* Texas A&M University, 2008.

Graves, A. W., Brandon, R., Duesbery, L., McIntosh, A., & Pyle, N. B. (2011). The effects of Tier 2 literacy instruction in sixth grade: Toward the development of a response-to-intervention model in middle school. *Learning Disability Quarterly, 34*(1), 73-86.

Harskamp, E. G., & Suhre, C. J. M. (2006). Improving mathematical problem solving: A computerized approach. *Computers in Human Behavior, 22*, 801-815.

Harskamp, E., & Suhre, C. (2007). Schoenfeld's problem solving theory in a student controlled learning environment. *Computers & Education, 49*, 822-839.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. Exceptional Children, 71, 165–179.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*(2), 99-107.

Hughes, R.A., Heron, J., Sterne, J.A.C., & Tilling, K. (2019). Accounting for missing data in statistical analyses: Multiple imputation is not always the answer. I*nternational Journal of Epidemiology, 48*(4), 1294-1304.

Joenssen, D.W., & Bankhofer, U. (2012). Donor limited hot deck imputation: Effects on parameter estimation. *Journal of Theoretical and Applied Computer Science, 6*(3), 58-70.

Kawalkar, A., & Vijapurkar, J. (2013). Scaffolding science talk: The role of teachers' questions in the inquiry classroom. *International Journal of Science Education, 35*(12), 2004-2027.

Kena, G., Aud, S., Johnson, F., Wang, X., Zhang, J., Rathbun, A., ... & Kristapovich, P. (2014). The Condition of Education 2014. NCES 2014-083. *National Center for Education Statistics*.

Kena, G., Hussar, W., McFarland, J., de Brey, C., Musu-Gillette, L., Wang, X., ... & Barmer, A. (2016). *The Condition of Education 2016* (NCES 2016-144). U.S. Department of Education, National Center for Education Statistics. Washington, DC.

Krippendorf, K. (1980). *Content analysis: An introduction to its methodology* (2nd ed., Chapter 11). Beverly Hills, CA: Sage.

Kuo, F. R., Hwang, G. J., & Lee, C. C. (2012). A hybrid approach to promoting students' web-based problem-solving competence and learning attitude. *Computers & Education, 58*(1), 351-364. doi: 10.1016/j.compedu.2011.09.020.

Lawrence, J. F., Crosson, A. C., Pare-Blagoev, E. J., & Snow, C. E. (2015). Word generation randomized trial: Discussion mediates the impact of program treatment on academic word learning. *American Educational Research Journal, 52*(4), 750-786. DOI: 10.3102/0002831215579485.

Lee, O., & Fradd, S. H. (1998). Science for all, including students from non-English-language backgrounds. *Educational Researcher, 27*(4), 12-21.

Lee, O., Maerten-Rivera, J., Penfield, R. D., LeRoy, K., & Secada, W. G. (2008). Science achievement of English Language Learners in urban elementary schools: Results of a first-year professional development intervention. *Journal of Research in Science Teaching, 45*(1), 31-52.

Lesaux, N. K., Kieffer, M. J., Kelley, J. G., & Harris, J. R. (2014). Effects of academic vocabulary instruction for linguistically diverse adolescents: Evidence from a randomized field trial. *American Educational Research Journal, 51*(6), 1159-1194. DOI: 10.3102/0002831214532165.

Li-Grining, C. P., Votruba-Drzal, E., Maldonado-Carreno, C., & Haas, K. (2010). Children’s early approaches to learning and academic trajectories through fifth grade. *Developmental Psychology, 46*(5), 1062-1077.

Manset-Williamson, G. & Nelson, J. M. (2005). Balanced, strategic reading instruction for upper-elementary and middle school students with reading disabilities: A comparative study of two approaches. *Learning Disability Quarterly, 28*, 59-74.

Mathan, S. A., & Koedinger, K. R. (2005). Fostering the intelligent novice: Learning from errors with metacognitive tutoring. *Educational Psychologist, 40*(4), 257-265.

Meyer, X. S., & Crawford, B. A. (2015). Multicultural inquiry toward demystifying scientific culture and learning science. *Science Education, 99*(4), 617-637.

Moore, D. S., Notz, W. I., Fligner, M. A. (2015). *The basic practice of statistics.* Seventh Ed. W. H. Freeman, New York, NY.

Myers, T. A. (2011). Hotdeck: An SPSS tool for handling missing data. *Communication Methods and Measures, 5*(4), 397-310.

Nelson, J. R., Vadasy, P. F., & Sanders, E. A. (2011). Efficacy of a Tier 2 supplemental root word vocabulary and decoding intervention with kindergarten Spanish-speaking English learners. *Journal of Literacy Research, 43*(2), 184-211.

Next Generation Science Standards (n.d.). *Next Generation Science Standards. Read the Standards.* Retrieved on March 7, 2018, from https://www.nextgenscience.org/sites/default/files/AllDCI.pdf

O'Connor, R. E., Gutierrez, G., Teague, K., Checca, C., Sun Kim, J., & Ho, T. (2013). Variations in practice reading aloud: Ten versus twenty minutes. *Scientific Studies of Reading, 17*(2), 134-162. DOI: 10.1080/10888438.2011.624566.

O'Connor, R. E., White, A., & Swanson, H. L. (2007). Repeated reading versus continuous reading: Influences on reading fluency and comprehension. *Exceptional Children, 74*(1), 31-46.

Pine, J., Aschbacher, P., Roth, E., Jones, M., McPhee, C., Martin, C., . . . Foley, B. (2006). Fifth graders' science inquiry abilities: A comparative study of students in hands-on and textbook curricula. *Journal of Research in Science Teaching, 43*(5), 467-484. DOI: 10.1002/tea.20140.

Pol, H. J., Harskamp, E. G., Suhre, C. J. M., & Goedhart, M. J. (2009). How indirect supportive digital help during and after solving physics problems can improve problem-solving abilities. *Computers & Education, 53*(1), 34-50.

Reardon, S. F., & Galindo, C. (2009). The Hispanic-White achievement gap in math and reading in the elementary grades. *American Educational Research Journal, 46*(3), 853-891.

Reiser, B. J., Tabak, I., Sandoval, W. A., Smith, B. K., Steinmuller, F., & Leone, A. J. (2001). BGuILE: Strategic and conceptual scaffolds for scientific inquiry in biology classrooms. In S. M. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 263-305). Mahwah, NJ: Erlbaum.

Santau, A. O., Maerten-Rivera, J. L., & Corinne Huggins, A. (2011). Science achievement of English language learners in urban elementary schools: Fourth-grade students achievement results from a professional development intervention. *Science Education, 95*, 771–793.

Saye, J. & Brush, T. (2001). The use of embedded scaffolds with hypermedia-supported student-centered learning. *Journal of Educational Multimedia and Hypermedia, 10*(4), 333-356. Norfolk, VA: Association for the Advancement of Computing in Education (AACE).

Saye, J. W. & Brush, T. (2002). Scaffolding critical reasoning about history and social issues in multimedia-supported learning environments. *Educational Technology Research and Development, 50*(3), 77-96.

Schmidt, H. G., Rotgans, J. I., & Yew, E. HJ. (2011). The process of problem-based learning: What works and why. *Medical Education, 45*, 792-806.

Sibold, C. Building English language learners' academic vocabulary: Strategies & tips. *Multicultural Education, suppl. Special Issue: Linguistically Diverse Students & Their Families, 18*(2), 24-28.

Silva, M., & Cain, K. (2014). The relations between lower and higher level comprehension skills and their role in prediction of early reading comprehension. *Journal of Educational Psychology, 107*(2), 321-331.

Silverman, R. D., Proctor, C. P., Harring, J. R., Doyle, B., Mitchell, M. A., & Meyer, A. G. (2013). Teachers' instruction and students' vocabulary comprehension: An exploratory study with English monolingual and Spanish-English bilingual students in grades 3-5. *Reading Research Quarterly, 49*(1), 31-60. DOI: 10.1002/rrq.63.

Simons, K. D., & Klein, J. D. (2007). The impact of scaffolding and student achievement levels in a problem-based learning environment. *Instructional Science, 35,* 41-72.

Snyder, T. D., de Brey, C., & Dillow, S. A. (2016). *Digest of Education Statistics 2014* (NCES 2016-006). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Snyder, T. D., de Brey, C., & Dillow, S. A. (2016). *Digest of Education Statistics 2015* (NCES 2016-014). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Snyder, T.D., de Brey, C., and Dillow, S.A. (2018). *Digest of Education Statistics 2016* (NCES 2017-094). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Stone, C. A. (1989). Improving the effectiveness of strategy training for learning disabled students: The role of communicational dynamics. *Remedial and Special Education, 10*(1), 35-42.

Sullivan Palincsar, A. (1986). The role of dialogue in providing scaffolded instruction. *Educational Psychologist, 21*(1 & 2), 73-98.

Tobin, K., Briscoe, C., & Holman, J. R. (1990). Overcoming constraints to effective elementary science teaching. *Science Teacher Education, 74*(4), 409-420.

Vadasy, P. F., Sanders, E. A., & Nelson, J. R. (2015). Effectiveness of supplemental Kindergarten vocabulary instruction for English learners: A randomized study of immediate and longer-term effects of two approaches. *Journal of Research on Educational Effectiveness, 8*(4), 490-529. DOI: 10.1080/19345747.2015.1033120.

Vygotsky, L. S. (1962). *Thought and language.* Cambridge, MA: MIT Press.

Vygotsky, L. S. (1978). In M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (Eds.), *Mind in society: The development of higher psychological processes.* Cambridge, *MA:* Harvard University Press.

Wood, P., Bruner, J., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17,* 89-100.

York-Barr, J., Ghere, G., & Sommerness, J. (2007). Collaborative teaching to increase ELL student learning: A three-year urban elementary case study. *Journal of Education for Students Placed at Risk (JESPAR), 12*(3), 301-335. DOI: 10.1080/10824660701601290.

APPENDICES

Appendix A

The Common Labs MC Vocab Test – 5th grade

The Common Labs MC Vocab Test – 5th grade

1. Capability

a) Having the knowledge to do a task.
b) Being limited in doing a task.
c) Having the potential for a known use.
d) Wearing a head cover with the favorite team’s name.

2. Fertilizer

a) Dressing used to make some salads tastier.
b) Things used to add minerals to the soil.
c) Food used specifically for animals.
d) Cover used to protect plants.

3. Versatile

a) Unable to finish a task.
b) Doing a task easily because of the skills.
c) Having many different uses.
d) A row in a poem.

4. Hydroponics

a) Training ponies near the water.
b) Growing plants without the support of soil.
c) Creating electric power with the help of water.
d) Watering plants with a solution of nutrients.

5. Dietary staple

a) Indulge yourself in tasty but healthy food.
b) Keeping track of weight loss meals.
c) Keeping diet recipes together easily.
d) Food that is eaten very often in people’s meals.

6. Supplies

a) Things used to torture people in the past.
b) Things you take away from someone who doesn’t need them.
c) Things you give to someone else that they need.
d) People who pray every night.

7. Facility

a) A place or building that is used for work in industry.
b) Work that is easy to do.
c) The ability of learning new things faster.
d) A service that you need to give for being paid.

8. Nutrient

a) Exercise that makes you stronger.
b) Food you store in cold places.
c) Food that makes you sick.
d) Things that help plants and animals grow.

9. Supplement

a) Something added to improve or make up for something missing.
b) Having an advantage over a weaker competitor.
c) Thing that is the base for building on it.
d) Thing that is superior compared to others.

10. Chamber

a) Loud noises that echo.
b) An enclosed space or section of a space.
c) A long walking pathway between buildings
d) A special type of transportation.

11. Moistened

a) Something covered in a thin layer of moss.
b) Something dried at high temperature.
c) Something that is damp or barely wet.
d) Something blessed by the pope.

12. Dissolve

a) To mix something in liquid until it disappears.
b) To use high temperature to melt something frozen.
c) To lose trust in something.
d) To make melted things or liquids solid.

13. Gravity

a) The capability of floating above ground.
b) The law of physics describing how Earth was formed.
c) The force of attraction that keeps all objects on Earth.
d) Being overweight when space travelling.

14. Solution

a) A mixture where one thing settles at the bottom.
b) A combination of elements used to build concrete.
c) A gesture through which people greet each other.
d) A uniform mixture of two or more things.

15. Colonize

a) To migrate to a place and settle in.
b) To conquer a difficult task for the first time.
c) To hold your ground in spite of difficulties.
d) To transplant an organ to save someone's life.

Appendix B

Classroom Observation Tool – Rubric and Work Sheet

for the 5th grade Science Module

Science Classroom Observation Tool – 5th Grade

Timing and Number of Occurrences of Instructional Features

	Start time	**Stop time**	**No. of occurrences**
Whole class instruction			
Independent seat work			
Group activities			
Technology activity (may be whole class, group, or individual)			
Integration of science with math			
Integration of science with English/Language Arts			

Science Classroom Observation Rubric – 5th Grade

No.	Feature	6	4	2	0
1	Alignment of lesson activities	Lesson activities directly addressed the stated learning objectives. It was very clear how the lesson activities would lead to a deeper student understanding of the learning objectives.	Lesson activities addressed the stated learning objectives but there was some question about how the lesson activities would lead to a deeper student understanding of the learning objectives.	Lesson activities addressed the stated learning objectives to some extent or poorly. It was difficult to understand how the lesson activities would lead to deeper student understanding of the learning objectives.	Lesson activities did not address the stated learning objectives. There was a clear mismatch between the stated learning objectives and the lesson activities.
2	Scaffolding (bringing out students' understanding of science concepts taught)	Most students articulated their current understanding of the science content. Most students were able to state for example why some radishes grew larger or a	Some students articulated their current understanding of the science content. Some students were able to state for example why some radishes grew larger or a	A few students articulated their current understanding of the science content. Only a few students were able to state for example why some radishes grew larger	Students did not articulate their current understanding of the science content. Very few or no students were able to state for example why some radishes grew larger

		larger percent of them survived, etc.	larger percent of them survived.	or a larger percent of them survived.	or a larger percent of them survived, etc.
3	Intellectual engagement	Most of the students were intellectually engaged with the science content related to the lesson activities. The learning tasks challenged most students to think at high cognitive levels, as observed in their group discussions.	Some of the students were intellectually engaged with the science content related to the lesson activities. The lesson challenged some students to think at high cognitive levels, as observed in their group discussions.	A few of the students were intellectually engaged with the science content related to the lesson activities. The lesson challenged a few students to think at high cognitive levels, as observed in their group discussions.	Students were generally intellectually unengaged with the science content related to the lesson activities, no students were challenged to use their critical thinking, as observed in their group discussions.
4	Use of evidence	Most students used evidence to explain their reasoning, back up their claims, or critique claims made by others, as observed in their responses to their classmates' questions or discussions.	Some students used evidence to explain their reasoning, back up their claims, or critique claims made by others, as observed in their responses to their classmates' questions or discussions.	A few students used evidence to explain their reasoning, back up their claims, or critique claims made by others, as observed in their responses to their classmates' questions or discussions.	Students did not have any opportunities to use evidence to explain their reasoning, back up their claims, or critique claims made by others, as observed in their responses to their classmates' questions or discussions.
5	P-BL (application of science)	Most of the students applied what they learned in the lesson to a new context – e.g. their experiment, or in a group activity, or in a written	Some students applied some things they learned in the lesson to a new context. Some of the students can explain (verbally to teacher or in	A few students applied something they learned in the lesson to a new context (their experiments, etc.). Only a few students can explain	There was no opportunity for students to apply something they learned in the lesson to a new context. Very few or no students can explain

		assignment, etc. Most students can explain (verbally to teacher or in small groups) how the concepts learned apply to planning, building their gardens, collaborating with NASA; also, making connections to what happens in the "real world" outside school, as observed in their classroom/group discussions.	small groups) how the concepts learned apply to planning, building their gardens, collaborating with NASA; also, making connections to what happens in the "real world" outside school, as observed in their classroom/group discussions.	(verbally to teacher or in small groups) how the concepts learned apply to planning, building their gardens, collaborating with NASA; also, making connections to what happens in the "real world" outside school, as observed in their classroom/group discussions.	(verbally to teacher or in small groups) how the concepts learned apply to planning, building their gardens, collaborating with NASA; also, making connections to what happens in the "real world" outside school, as observed in their classroom/group discussions.
6	Science and Engineering practices (SEP) & integration w. math	Most of the students were able to plan their garden in their small groups or individually, to explain what the drawing means (drawing, labels, explanations are all required by NGSS), to build it, to make and record their observations, to interpret data correctly, or to answer questions in writing or verbally about the teacher's lecture and presenting the experiment to the class, and to interpret data from teacher's experiment, as observed in their classroom activities.	Some of the students were able to plan their garden in their small groups, to explain what the drawing means (drawing, labels, explanations are all required by NGSS), to build it, to make and record their observations, to interpret data correctly, or to answer questions in writing or verbally about the teacher's lecture and presenting the experiment to the class, and to interpret data from teacher's experiment, as observed in their classroom activities.	Few students were able to plan their garden in their small groups, to explain what the drawing means (drawing, labels, explanations are all required by NGSS), to build it, to make and record their observations, to interpret data correctly, or to answer questions in writing or verbally about the teacher's lecture and presenting the experiment to the class, and to interpret data from teacher's experiment, as observed in their classroom activities.	Very few or none of the students were able to plan their garden in their small groups, to explain what the drawing means (drawing, labels, explanations are all required by NGSS), to build it, to make and record their observations, to interpret data correctly, or to answer questions in writing or verbally about the teacher's lecture and presenting the experiment to the class, and to interpret data from teacher's experiment, as observed in their classroom activities.

7	Formative assessment	The teacher continually assessed the depth of student understanding of the learning objectives, and when appropriate, adjusted instruction accordingly, providing extra support to the ELLs and students with special needs.	The teacher occasionally assessed the depth of student understanding of the learning objectives, and when appropriate, adjusted instruction accordingly, providing some extra support to the ELLs and students with special needs.	The teacher rarely assessed the depth of student understanding of the learning objectives, and when appropriate, adjusted instruction accordingly, providing little support to the ELLs and students with special needs.	There was little or no evidence that the teacher assessed the depth of student understanding of the learning objectives, providing little or no extra support to the ELLs and students with special needs.
8	Group work	Most students were actively involved in their group work regarding the academic discussions of the content taught, then planning of the task at hand, then completing their task.	Some students were involved in their group's work involving academic discussions of the content taught, then planning of the task assigned, and completing their task.	Few students were involved in their group's work involving academic discussions of the content taught, then planning of the task at hand, and completing their task.	Students were not actively involved in their group's work when discussing the content taught, then planning the task assigned, or they were working in different groups than assigned.
9	Teacher scaffolding – students with special needs	Teacher provided scaffolded support to the students with special needs, when needed. During their initial reading of the lesson materials, the students with disabilities were supported by teacher (T) or peers (P).	Teacher provided scaffolding scarcely during the lesson to the students with disabilities, only a few scaffolding instances, less than needed. Some (but not all) of the students with special needs were specifically receiving scaffolded teacher (T) or peer (P) support.	Teacher provided hardly any scaffolding during the lesson: almost no individual scaffolding instances, way less than needed and/or requested. Few of the students with special needs were specifically receiving scaffolded teacher (T) or peer (P) support.	Teacher provided no scaffolding during the lesson: almost no individual scaffolding instances, or none at all to the students with disabilities. Very few or none of the students with special needs were specifically receiving scaffolded teacher (T) or peer (P) support.

10	Teacher scaffolding – students who are ELLs	Teacher provided scaffolded support to the students who are ELLs when needed. During their initial reading of the lesson materials, the ELLs were supported by the teacher (T) or peers (P).	Teacher provided scaffolding scarcely during the lesson to the students who are ELLs, only a few scaffolding instances, less than needed. Some (but not all) of the ELLs were specifically receiving scaffolded teacher (T) or peer (P) support.	Teacher provided hardly any scaffolding during the lesson: almost no individual scaffolding instances, way less than needed and/or requested. Few of the students who are ELLs were specifically receiving scaffolded teacher (T) or peer (P) support.	Teacher provided no scaffolding during the lesson: almost no individual scaffolding instances, or none at all to the students who are ELLs. Very few or none of the students who are ELLs were specifically receiving scaffolded teacher (T) or peer (P) support.
11	Enthusiasm – students with special needs	Instruction was presented in an accurate, meaningful, memorable, and engaging manner with specific support for the students with special needs.	Instruction was presented in a moderately accurate, meaningful, memorable, and engaging manner with some support for the students with special needs.	Instruction was presented in a slightly accurate, meaningful, memorable, and engaging manner with little support for the students with special needs.	Instruction was not accurately presented, nor was it meaningful, memorable, or engaging with very little or no support for the students with special needs.
12	Enthusiasm – students who are ELLs	Instruction was presented in an accurate, meaningful, memorable, and engaging manner with specific support for the ELLs.	Instruction was presented in a moderately accurate, meaningful, memorable, and engaging manner with some support for the ELLs.	Instruction was presented in a slightly accurate, meaningful, memorable, and engaging manner with little support for ELLs.	Instruction was not accurately presented, nor was it meaningful, memorable, or engaging with very little or no support for the ELLs.
13	Rapport between teacher and students	For the entire class, the teacher-students relationship was mutually positive and accepting.	For most of the class or a majority of the students, the teacher-students relationship was mutually positive and accepting.	For a short time during the class or for a very few students, the teacher-student relationship was mutually positive and accepting.	The teacher-students relationship is not mutually positive and accepting, or a teacher-student rapport is completely absent.

From: RMC Research Corporation u Portland, Oregon
and: LASER – Leadership and Assistance for Science Education Reform

and: DBR-CM ER – Dr. Sims, GSoE, U.C. Riverside
rubric items added and modified to fit the purpose of the study by Mima Laptes-Frangu, GSoE, U.C. Riverside

Science Classroom Observation Worksheet

School & District:__________________________________ **Date**: ______________________

Teacher:____________________________________ **Grade/Subject**: _____________________

Observer: ____

Rationale for Rating	**Rating**
1. Alignment of lesson activities	
	0 1 2 3 4 5 6
2. Scaffolding (bringing out students' understanding of science concepts taught)	
	0 1 2 3 4 5 6

Rationale for Rating **Rating**

3. Intellectual engagement

0 1 2 3 4 5 6

4. Use of evidence

0 1 2 3 4 5 6

5. P-BL (appplication of science)

0 1 2 3 4 5 6

6. Science and engineering practices & integration w. mathematics

0 1 2 3 4 5 6

Rationale for Rating	**Rating**

7. Formative assessment

0 1 2 3 4 5 6

8. Group work

0 1 2 3 4 5 6

9. Teacher scaffolding – students with special needs

0 1 2 3 4 5 6

10. Teacher scaffolding – students who are ELLs

0 1 2 3 4 5 6

Rationale for Rating	**Rating**

11. Enthusiasm – students with special needs

0 1 2 3 4 5 6

12. Enthusiasm – students who are ELLs

0 1 2 3 4 5 6

13. Rapport between teacher and students

0 1 2 3 4 5 6