

Article

Planning Science Instruction for Critical Thinking: Two Urban Elementary Teachers' Responses to a State Science Assessment

Elaine Silva Mangiante

School of Education, University of Rhode Island, Kingston, RI 02881, USA;

E-Mails: emangiante@my.uri.edu; emangiante@gmail.com; Tel.: +1-401-874-4068;

Fax: +1-401-874-5471

Received: 6 February 2013; in revised form: 17 June 2013 / Accepted: 19 June 2013 /

Published: 8 July 2013

Abstract: Science education reform standards have shifted focus from exploration and experimentation to evidence-based explanation and argumentation to prepare students with knowledge for a changing workforce and critical thinking skills to evaluate issues requiring increasing scientific literacy. However, in urban schools serving poor, diverse populations, where the priority is on students' assessment results in reading and math, students may not receive reform-based science. The rationale for this qualitative study was to examine how two elementary teachers from high-poverty urban schools planned for reform-based science in response to a quality state science assessment in conjunction with their training and resources. Their state assessment included an inquiry task requiring students to construct responses to questions based on their investigation data. From evaluating evidence using Zembal-Saul's continuum for teaching science as argument, the findings indicated that both teachers adopted an investigation-based and evidence-based approach to science teaching to prepare students for the inquiry task. However, one teacher provided argument-based science teaching from her explicit training in that approach. The results suggested that the teachers' training and resources informed their interpretation of the focus areas on the science assessment inquiry task and influenced the extent to which they offered students an equitable opportunity to develop higher-order thinking from reform-based science.

Keywords: elementary science; urban schools; assessments; critical thinking; teacher planning; argumentation; reform-based science

1. Introduction

Educational systems in the twenty-first century face challenges in preparing all students with knowledge needed for a rapidly changing technical and scientific workforce, as well as critical thinking skills to evaluate national and global issues requiring increasing scientific literacy [1]. To influence the practices of educators in science teaching, science education reform in the United States has been shaped by a two-pronged national initiative, including development of standards specifying “what students need to know, understand, and be able to do to be scientifically literate” ([2], p. 2) and assessments measuring students’ science learning outcomes [3].

For over a half century, the US has placed increased emphasis on science education reform. To increase the national welfare, security and competitiveness, the National Science Foundation was established in 1950 with the goals to cultivate the science and engineering workforce and expand the scientific knowledge of all citizens [4]. In an effort to promote these goals in education, the National Academy of Sciences began curriculum development efforts in 1959 advocating the “discovery method” by which students engage in “hands-on learning” for science and math ([5], pp. 33–34). To guide teachers in implementing this new process learning approach, a three-phase learning cycle of exploration, invention and discovery was introduced for elementary teachers in the 1960s [6] and the 5E Instructional Model (engage, explore, explain, extend, evaluate) in the 1980s [7]. Yet, this reform focus on discovery and exploration shifted with the National Research Council’s establishment of science standards outlining content and inquiry process skills necessary for students’ scientific literacy [2]. As science education reform has evolved [1,2,8,9], the next generation of science standards has emphasized students’ higher-order thinking through practices of reasoning, problem solving, discourse and debate [10].

The second prong of the US national initiative, assessment of student learning, began in 2001 with the federal authorization of Title 1 of the Elementary and Secondary Act, commonly referred to as No Child Left Behind (NCLB), an accountability approach for measuring student’s academic achievement [3]. Under this act, state education departments are required to establish academic standards, assessments and accountability systems to ensure that students in all public elementary and secondary schools make adequate yearly progress (AYP) in the subject areas of mathematics and reading/language arts [3]. In addition, AYP is determined by the separate measure of achievement by different student groups, including economically disadvantaged students, students from major racial and ethnic groups, students with disabilities and students with limited English proficiency. An implication of this requirement is that for school districts with more student groups, such as urban schools, demonstrating AYP can be more challenging [13]. With regard to the subject area of science, measurement of US students’ science learning began in 2008; however, the scores are not used to calculate a school’s annual yearly progress (AYP) [3].

In urban elementary schools serving economically disadvantaged and racially/ethnically diverse populations, where accountability pressure is focused in the high stakes subjects of mathematics and reading, studies have shown that allocation of instructional time and resources for science is often not a priority and teachers’ implementation of reform-based science is de-emphasized [11–13]. Research has also indicated that when state science assessments measure factual knowledge through multiple-choice formats, teachers’ instruction typically involves rote learning to prepare students for the test, rather

than the intent of the standards for higher-order thinking [14–18]. These factors serve as barriers to teachers' adoption of reform-based science, particularly in schools that struggle to meet AYP goals. Thus, students in the poorest and neediest schools may be denied an equitable opportunity to acquire critical thinking skills for scientific literacy.

Addressing the alignment of assessment design to the standards, Darling-Hammond asserted that high-quality assessments should “emphasize deep knowledge of core concepts within and across the disciplines, problem solving, collaboration, analysis, synthesis, and critical thinking” ([19], p. 3) incorporating “more analytic selected-response and open-ended items than many U.S. tests currently include” ([19], p. 8). She cited the New England Common Assessment Program (NECAP) as an assessment system designed to improve the quality of students' learning, rather than only to measure it.

This research emerged based on data collected as part of a larger study examining the beliefs, knowledge bases and resources impacting the science reform planning of two fourth grade teachers in high-poverty urban schools. Unexpected reports by both teachers indicated that the science test used by their state, the NECAP, had influenced the shift in their science teaching from rote textbook learning to reform-based science. Despite the pressure they experienced to increase student scores in math and reading, they maintained time for science instruction focused on their students' critical thinking development.

The rationale for this qualitative study was to examine this phenomenon more deeply to explain how these teachers planned for science instruction in response to the science test and how they integrated their understanding of the state science test with the view of reform-based science they developed from their training and resources. For this study, the term, *reform-based science*, is used to convey the current standards' intent for students' critical thinking development [10]. The warrant for the research was to provide an in-depth description of what was *possible* in the teachers' planning to support students' equitable learning for scientific literacy in districts that used high-quality science assessments, despite obstacles in the urban context. The research questions included:

- (1) How do two urban fourth grade teachers plan for reform-based science instruction in response to the format of the state science assessment?
- (2) How do the teachers integrate their understanding of the state science assessment with resources and training they have available to plan for reform-based science instruction?

2. Literature Review

The literature that informed this study was drawn from four areas: the role of education in critical thinking development, standards-based scientific practices to develop students' scientific literacy, effective equitable pedagogy in science by urban elementary teachers and formats of elementary science assessments coherent with reform standards.

2.1. The Role of Education in Critical Thinking Development

The capacity for citizens to make informed decisions in a society involves the skill of critical thinking. Kuhn [20], a researcher in the area of cognitive development, studied the nature and acquisition of critical thinking skills and the role that education plays in its development. She posited

that “developing the competencies that enable people to participate fully as citizens in a democracy remains the unifying purpose, and great promise, of public education” ([20], p. 6). To depict ways of knowing, Kuhn [20] defined categories of epistemological thinking: (a) realist—knowledge is certain, coming from an external source; (b) absolutist—knowledge is certain, coming from an external source, but assertions can be correct or incorrect; (c) multiplist—knowledge is uncertain, because it is generated by equally valid human opinions; and (d) evaluativist—knowledge is uncertain and evaluated according to criteria using evidence. Based on this framework, Kuhn and Weinstock argued that a person who approaches knowledge acquisition as an evaluativist has the competency to think critically by “judging some claims as having more merit than others” rather than viewing knowledge as certain, coming from an external source or from one’s personal opinion ([21], p. 126).

Brown [22] attested that students can internalize critical thinking skills for problem-solving when these skills are modeled by teachers, suggesting that teachers’ epistemological approach and instructional choices can make a difference to students’ critical thinking development. However, research has indicated that teachers have strongly held epistemological views of science as a body of scientific facts to be transmitted to students [23]. In studies examining the science teaching practices of elementary teachers in urban schools, teachers described themselves as facilitators of students’ critical thinking and inquiry science learning; yet, their practice was more didactic and expository in nature [24–26]. Elementary teachers often fear the noise, the mess or potential conflicts when students work in small groups for science [24,27]. These fears and their underlying epistemological belief in the transmission of scientific information can result in teachers controlling the procedure of science instruction through sequenced steps that lead students to the “right answer” without giving students the opportunity for higher-level thinking ([24], p. 848). These factors are compounded in urban schools if teachers hold deficit beliefs about their students; a mindset that can impede reform implementation and equitable learning opportunities for poor, diverse student populations [28]. Delpit [29] stressed the importance of empowering marginalized students with tools for success. Kuhn’s [20] emphasis on the value of promoting students’ ability to think and critically evaluate assertions can serve all students as future citizens. Yet, teachers’ competency and explicit planning for critical thinking can impact the critical thinking skill development of their students [22].

2.2. Standards-Based Scientific Practices to Develop Students’ Scientific Literacy

The US National Science Education Standards (NSES), developed by the National Research Council (NRC) [2], established a vision for science education that students develop higher-order thinking skills for scientific literacy when confronted with questions that require scientific information and analysis. To accomplish this goal, the NSES recommended that students actively learn through “inquiry,” defined as a “set of interrelated processes by which...students pose questions about the natural world and investigate phenomena” in order to acquire knowledge and develop an understanding of scientific concepts and principles ([2], p. 214). The document, *Inquiry and the National Science Education Standards*, further clarified that inquiry involved five essential features:

- Learners are engaged by scientifically-oriented questions.
- Learners give priority to evidence, which allows them to develop and evaluate explanations that address scientifically-oriented questions.

- Learners formulate explanations from evidence to address scientifically-oriented questions.
- Learners evaluate their explanations in light of alternative explanations, particularly, those reflecting scientific understanding.
- Learners communicate and justify their proposed explanations. ([8], p. 25, *emphasis in original text*)

The goal of science education had shifted from the 1960s focus on the process skills of exploration and experimentation [30] to the use of evidence to explain phenomenon with an emphasis on “science as argument and explanation” ([2], p. 113). The distinction between these two scientific practices in the K–8 classroom is that students develop *explanations* by making a claim supported by available evidence [31,32]; whereas, students add to explanation construction by engaging in *argumentation* dialogue through evaluating evidence, persuading peers of the logic of an explanation, responding to critiques, debating alternative explanations and negotiating consensus [33,34].

The document, *Science for All*, conveyed the need for reform in science education: “Students cannot learn to think critically, analyze information, communicate scientific ideas, make logical arguments, work as part of a team,... unless they are permitted and encouraged to do those things” ([35], p. 187). To aid teachers in developing students’ inquiry skills, NRC presented a continuum of pedagogical variations from “guided” to “open” inquiry ([8], p. 29). At first, students learn to use evidence to formulate explanations from teacher-provided questions, materials and step-by-step procedures. With continued guidance, students take on more responsibility for investigations and meaning making. Through open inquiry, students take ownership of the focus question, experimental design, data collection/recording, explanation building and logical argument to justify explanations.

However, reviews of science education policy and meta-analyses of studies examining the implementation of inquiry-based pedagogy indicated that the term “inquiry” had been interpreted in different ways by practitioners and researchers [36,37]. Thus, NRC [9] developed a framework on which to base the next generation of K–12 science education standards describing the scientific and engineering practices all students should acquire, which include:

- Asking questions (for science) and defining problems (for engineering)
- Developing and using models
- Planning and carrying out investigations
- Analyzing and interpreting data
- Using mathematics and computational thinking
- Constructing explanations (for science) and designing solutions (for engineering)
- Engaging in argument from evidence
- Obtaining, evaluating and communicating information ([9], p. 49)

The framework called for integration of these practices with science content rather than focusing on procedures, recognizing teachers’ tendency to overemphasize science investigation as a step-by-step procedure at the expense of essential practices for higher-order thinking of modeling, explanation, argumentation and communication [9,23].

To emphasize the focus on explanation and argument, Zembal-Saul [38] advanced a *continuum for teaching science as argument* framework representing four approaches to teaching science with increasing alignment to the science standards:

- (1) Activity-based—fun, hands-on activities designed to motivate students and keep them physically engaged
- (2) Investigation-based—abilities to engage in inquiry [2], ask testable questions and design fair tests; focus on collecting data
- (3) Evidence-based—need to support claims with evidence; evidence is not questioned in terms of quality, coherence, *etc.*
- (4) Argument-based—argument construction is central, coordinating evidence and claims is viewed as important; emerging attention to considering alternatives. ([38], p. 703)

To shift pre-service teachers' science education pedagogy beyond the activity- and investigation-based levels, Zembal-Saul [38] used this framework during an elementary science methods course to assist future teachers in promoting students' sense-making and making their thinking visible. Zembal-Saul found that for those future teachers identified to be thinking at the activity-based level, they regarded science instruction as involving fun, hands-on activities. Those at the investigation-based level developed an awareness of the abilities students would need to do science, such as asking testable questions and devising fair tests with the goal to produce data. However, the intent was to summarize the collected data as the final outcome of the investigation. Some pre-service teachers at the evidence-based level recognized the importance of students' voicing their claims based on evidence during class discussions; yet, they did not consider the quality of the evidence or the plausibility of the claim; thus, regarding the mere citing of evidence for a claim as adequate. For the argument-based level, few future teachers made reference to students' evaluating each other's results to develop an explanation or considering additional data collection in order to generate an alternative claim.

The results from research using the continuum indicated that it was possible for pre-service teachers to advance from the activity-based view of science teaching to the investigation-based approach [38]. However, it was not as likely for them to adopt an evidence- or argument-based approach. These findings established an entry point for Zembal-Saul to promote teachers' focus on discourse with their students in identifying patterns from the data to generate claims, constructing evidence-based explanations, evaluating competing claims from evidence, considering alternative explanations and ascertaining if additional evidence is needed. This approach was supported by research indicating that by creating a classroom climate of making scientific reasoning public, teachers and students could engage in the process of critical thinking [39]. The continuum was adopted for this study as the conceptual framework for analyzing the science teaching practice of in-service elementary teachers.

2.3. Effective Equitable Science Pedagogy by Urban Elementary Teachers

Science reform practices pose challenges for many teachers in urban schools. These practices conflict with the commonly held deficit view of poor, diverse students and the pedagogical approach used by urban teachers which "cast students as passive recipients of 'basic' science facts" ([40], pp. 21–22). In addition, elementary teachers generally have limited knowledge of science reform pedagogy and tend to offer 'hands-on activities' for science instruction without encouraging students "to think deeply about the phenomena being explored" ([41], p. 1169) or promoting discourse and argumentation [42]. The teachers' lack of skill in argumentation and view of science as "an

unproblematic collation of facts” can result in students’ conclusions from investigations being unquestioned and their capacity to critique scientific claims remaining undeveloped ([33], p. 288). Driver *et al.* [33] argued that teachers must make discourse explicit by establishing norms of scientific argumentation through comparing claims and considering alternative hypotheses.

However, research has indicated examples of urban teachers who have countered the deficit paradigm and succeeded in providing reform-based science teaching in urban elementary schools. Varelas and colleagues [43] identified urban teachers who implemented pedagogical practices to nurture early elementary students’ explanation-making and argumentation in science by challenging their students’ thinking, expecting students to “prove” or offer arguments for their claims and encouraging students to consider alternative ways of viewing phenomenon. Research has shown that effective teachers use high quality questioning targeted to student needs and “incremental build-up of skills and knowledge over time” to support students in “making connections for themselves” ([44], p. 145). During the “meaning making conference” of a science lesson, Amaral and Garrison reported that a fourth grade teacher, whose class included Spanish-speaking students, helped students make connections between their own investigation findings and scientific concepts by asking probing questions and expecting students to provide evidence for their claims ([45], p. 159). From research in urban schools, Carlone, Haun-Frank and Webb found that students can engage in scientific discourse and become “collaborative producers of knowledge”, building on and questioning other’s ideas when teachers establish equitable participation structures ([46], p. 480). Upadhyay [47] reported on a Hispanic teacher who believed that by connecting students’ science experiences to their lives, while also preparing them for the state science test, she could empower her students from disenfranchised communities to engage in science and keep the door open for their continued education.

2.4. Formats of Elementary Science Assessments Coherent with Reform Standards

When assessing students’ learning, the form of a standardized test can influence the information that teachers obtain [48], as well as the instructional decisions that teachers make [49,50]. The intent of standards-based reform policy is that the standards inform classroom instruction with assessment tools measuring the extent to which students have improved their learning based on the standards [15]; thus, shifting teacher practice to be in alignment with reforms [51,52]. However, studies have indicated that accountability tests are driving the teachers’ instructional decisions more than the standards [50], impacting the content taught, the form of the knowledge acquired and the pedagogical approaches used in the classroom [53].

A consequence of the NCLB federal testing mandate [3] has been increased pressure placed on schools serving poor, diverse populations with more student groups to meet AYP than less diverse suburban communities [54,55]. Teachers in urban schools have dealt with the pressure by teaching for test preparation and lower level knowledge [14,56] and focusing on the high-stakes subjects of reading and mathematics, sacrificing instructional time for science not counted toward AYP [11,57]. Marx and Harris posited that reform-based pedagogy would become “upper class science” in schools that could afford to devote time to critical thinking ([54], p. 471). This school differential in opportunities would be inequitable given evidence that students make learning gains from reform-based science [58,59].

Studies have indicated that multiple choice test formats tend to result in teachers' narrowing the taught curriculum and preparing students for single-item questions rather than for critical thinking [53,60,61]. Pedulla *et al.* [61] found it more prevalent for elementary teachers to narrow the curriculum and spend time on tested areas than secondary teachers. They speculated that with so many subjects to teach, the test provided elementary teachers with guidance on areas to cover. However, from a meta-analysis of 49 qualitative studies of 740 teachers, Au [53] reported that though high-stakes tests tended to result in teachers' narrowing the curriculum, certain test types promoted an expansion of curriculum content, integration of knowledge and student-centered pedagogies. This finding suggests that test design matters and raises the question of what tests would look like to promote higher-order thinking.

Research has provided recommendations for test formats that align to the reform intent. Liu, Lee, Hofstetter and Linn reported that explanation assessment items were more effective in differentiating students' science performance than multiple choice items [62]. They proposed test items that "reward complex thinking" skills rather than "focusing on discrete facts" ([62], p. 53). Yeh argued that as early as fourth grade, state-mandated tests should assess students' critical thinking in developing evidence-based claims and evaluating arguments rather than recall of rote factual learning, asserting that a test could combine "both open-ended and forced choice items...to assess critical thinking in a practical, cost-effective way" ([63], p. 16). From an examination of researcher-designed assessments to fuse science content with explanation-building for elementary students in high poverty schools, Songer and Wenk Gotwals suggested that more open-ended assessments prompting students for evidence could provide information about how students make the practice of explanation their own [64].

In summary, reform-based science standards emphasize higher-level thinking through students' explanation and argumentation, rather than a procedural focus on exploration and experimentation. Teachers' adoption of the intent of these standards in their instruction is impacted, in part, by the pressures they experience from the format and content of the state science test, particularly in urban schools. Research has highlighted the nature of high quality assessments that emphasize students' critical thinking [62], as well as practices of effective urban teachers who have provided science instruction with fidelity to the standards [43,47].

3. Methodology

This study used qualitative multi-case methodology employing naturalistic inquiry [65,66] and an interpretive approach [67] to examine the phenomenon of two urban elementary teachers' planning for reform-based science and the decisions they made in response to the state science assessment.

The participants in the study were selected from fourth grade teachers in urban districts in a northeastern state of the US with the highest percentage of children under age 18 living below the federal poverty threshold [68]. The selection process involved both nomination and observation. Other studies have used recommendations by fellow teachers, community members and school leaders and observations to identify effective teachers in urban schools [69,70] as a purposeful means to select study participants [67,71]. Nominations were based on two sets of criteria: (a) characteristics of teachers effectively serving students in urban schools [44,69,70,72–74] and (b) pedagogical practices of teachers providing reform-based science [1,8,9]. A drawback of utilizing nominations for participant

selection is that the nominators' judgment is subject to their knowledge of the teachers and their own expertise in science teaching. To address this limitation, curriculum coordinators, science content coaches and science education consultants were chosen who worked closely with the teachers and who had knowledge of their ability in teaching students in urban schools, as well as their skills with science reform pedagogy. In addition, the researcher, unaffiliated with the districts, observed lessons taught by nominees to confirm their effectiveness in the urban context and with reform-based science.

Two fourth grade teachers, Ann and Lee (pseudonyms), from different urban school districts were selected from a pool of eight nominated teachers. Nominators described Ann as a teacher who provided challenging and engaging science lessons and had confidence in students' abilities to meet her high expectations. Both the nominators' evaluations and the researcher's observation indicated that Ann was highly skilled at accessing student's prior knowledge, addressing teachable moments, aiding all students to succeed, questioning, expecting students to identify evidence from their investigation and providing opportunities to discuss the credibility of data collected, students' explanations and merits of each other's responses. Nominators for Lee explained that she had been developing her science content and pedagogical knowledge for the last few years. From evaluations of her science teaching practice by nominators and the researcher's observation, Lee received top ratings in her ability to provide opportunities for her students to achieve excellence, promote students' generation of testable questions and development of data recording systems, connect science learning to real life experiences and facilitate class dialogue in science.

3.1. Setting and Participants for the Study

Both Ann and Lee were of white racial background and had taught in their urban districts for 14 years. They had adopted a reform-based approach to science learning over the course of three to five years with professional development and mentorship support.

Ann taught fourth grade in an urban district where 40.9% of the children lived below the poverty level, the highest poverty area among the state's districts, with 100% of the students receiving free or reduced lunch [68]. Her class included 23 students: 19 of Hispanic descent from the Dominican Republic, Puerto Rico and Columbia; two white; one African American; and one Cape Verdean. Two students received English as a Second Language services, and six other students had learning differences. Based on information provided by the school administrator, the school met its improvement goals for the year prior to the study; however, it was still considered a "low performing" school on probation in terms of AYP status [3]; the school needed to meet AYP goals two years in a row in order to elevate its status from "Delay" to "Met AYP." Ann's school was assessed based on the performance of six different student groups of 45 students or more, which included all students, Hispanic students, white students, students with learning disabilities, English language learners and economically disadvantaged students. Though all student groups made sufficient progress in math and English/language arts for AYP, none of the groups met the target goal for proficiency.

Lee was a fourth grade teacher in an urban district where 25.3% of the children lived below the poverty level. This district was the fourth highest poverty area in the state, and 73% of the students received free or reduced lunch [68]. Her class was comprised of 26 students: seven of Hispanic or Cape Verdean descent, ten white and nine African American. Six students received special education

services, and eight other students were eligible for academic support. Out of three classes, Lee's classroom served fourth grade students with special needs. Based on information provided by the school administrator, Lee's school met AYP goals or made sufficient progress in the year prior to the study [3]. Lee's school was assessed based on the performance of six different student groups of 45 students or more, which included all students, African-American students, Hispanic students, white students, students with disabilities and economically disadvantaged students. For English/language arts, only the Hispanic student group met the target proficiency score for the year; yet, all other student groups made sufficient progress. In mathematics, all racial groups met the target score, while students with disabilities and economically disadvantaged students made sufficient progress.

3.2. Data Sources

Data sources included interviews, observations of planning meetings, documents and observations of lessons [67] collected over the course of the teachers' first science unit in the fall of 2011. These data collection procedures, typical for case studies, were in alignment with research methods used to study teacher planning [75]. The rationale for choosing the first science unit was based on findings that teachers' planning decisions made early in the year profoundly influence subsequent planning for the remainder of the year [76]. Since the focus of this study was on the cognitive process of teacher planning, interviewing was a means to discover the "feelings, thoughts, and intentions" of the teachers' thinking in preparing for science lessons ([71], p. 341). Guba and Lincoln [59] regard interviewing as indispensable in tapping into the experience of others. They describe advantages of this form of data collection in that it is likely to provide a more in-depth examination of a phenomenon and opportunities for the researcher to probe further or explore fruitful leads from evaluation of the participant's responses than other forms of inquiry [59]. Underlying qualitative interviewing is the assumption that that the perspective of a participant is "meaningful, knowable, and able to be made explicit" ([71], p. 341). To gain insight into each teacher's thinking when planning for science, data collection involved semi-structured interviews conducted using a general interview guide to ensure consistency in the questions explored with both teachers, while also allowing topics to emerge specific to each teacher's context [65,67,71] (see Appendix 1). The interviews were conducted in the naturalistic setting of the teachers' classrooms in order for teachers to have access to their planning materials and student work artifacts. The teachers participated in 45 minute audio-recorded interviews once a week during the science unit to share their planning for science lessons. Seven interviews were conducted with Ann and six with Lee. In addition, two of Ann's weekly planning meetings with a colleague were observed and recorded.

The teachers used documents in their planning for science lessons, including sample NECAP science tests released to teachers by the state department of education, lesson plan books, teacher-made worksheets, science kit teachers' guides and student worksheets and resources for science content. Emerging findings prompted the examination of these documents to triangulate information obtained from interviews or observations of planning meetings [67].

In addition, to account for planned routines that the teachers may have taken for granted, the researcher used direct naturalistic observations at the field site [71]. For this study, two science lessons conducted by Ann and Lee were observed to identify routine practices not mentioned by the teacher

that required planning and to generate questions about prior planning, as well as future planning. In alignment with the study's focus on teacher thinking, the observations served to reveal unreported planning decisions that were explored with the teachers during follow-up interviews; it was not to examine the implementation of the decisions.

Thus, to capture a more complete understanding of the teachers' planning for their reform-based science unit, Ann and Lee responded each week to the same key questions from the general interview guide to gain access to their planning decisions for that week. However, additional questions were generated immediately following each interview, document analysis, observation of a planning meeting or observation of a lesson to be asked during the next interview. For example, as the teachers referenced their understanding of the NECAP science test in planning for science lessons and the challenges they faced in their urban context, follow-up questions were shaped to target each teacher's view of her focus areas informed by the test and how she planned to achieve her goals (see Appendix 1). The follow-up, detail-oriented questions served to delve more deeply into each teacher's responses [71] for clarification and critical reflection [66].

3.3. Data Analysis

With a naturalistic inquiry approach, a two-fold iterative process was used for analysis at the research site and following data collection [65,67]. Insights emerged during the interviews and observations, as well as from the examined documents regarding the teachers' understanding of the science assessment and its impact on their planning. The interview transcriptions, notes from planning meetings, document notes and observation field notes for each case initially were coded based on the Zembal-Saul [38] continuum for teaching science (see Table 1). The codes were generated from the Zembal-Saul's four key modes of teaching science (activity, investigation, evidence and argument) and sub-codes representing properties of each, as defined by Zembal-Saul. Codes 2.1 and 2.3 were added to the list of codes to make a distinction between planning for general student questions *vs.* testable questions and planning for a general investigation design *vs.* a fair test. Specifically, each teacher's instructional decisions were coded along this continuum to generate the data corpus for deductive analysis of how Ann and Lee's planning decisions reflected Zembal-Saul's theory [71]. However, from insights emerging during the data collection and from a systematic and repeated search of the data corpus [77], an open coding system was adopted to examine the data anew for undiscovered understandings [71] (See Appendix 2). For example, both teachers described how they planned for the teacher and student role in the science classroom and to develop students' capacity in questioning, use of science language and written representations in response to the NECAP science test. Sorting the data through the lens of these emerging codes allowed the researcher to examine practices in greater depth and to generate themes from patterns in the teachers' planning [67]. Thus, the themes identified inductively from the data provided further explanation for each teacher's science planning that was categorized deductively along the Zembal-Saul continuum [38]. The themes from the two cases were examined for similarities and differences between the teachers regarding the questions under investigation in order to identify key findings for the study, as well as to explain variation between the teachers' planning within their context [66,77].

To enhance the trustworthiness of the analysis, the researcher engaged in separate weekly conversations with two peer debriefers, professional colleagues outside the context of the study [65]. As the researcher presented the new data and themes under consideration, each debriefer would ask probing questions, provide counterarguments, suggest additional interview questions and propose alternative explanations. For example, when considering interview questions focused on answering the first research question, one debriefer suggested rephrasing “How do you plan in response to the NECAP science test?” to “What steps do you take...” to increase the specificity of the question and to reduce participants’ concern about the possibility of “getting in my head or judging my thinking.” To inform final interpretations of the data for the second research question, one debriefer proposed that the training in reform-based science each teacher received may have emphasized investigational processes rather than explanation and argumentation.

Table 1. Initial codes and examples of planning decisions.

| Initial Codes | Examples from Ann | Examples from Lee |
|---|---|--|
| Plan for Activity | | |
| 1.1 Plan for fun hands-on activity to motivate students | “They loved it. The magnets, they love the magnets.” <i>But no data regarding planning for this sole purpose.</i> | “Science is so much fun for them, it’s so motivational.” <i>But no data regarding planning for this sole purpose.</i> |
| 1.2 Plan for activity to keep students physically engaged | <i>No data</i> | <i>No data</i> |
| Plan for Investigation | | |
| 2.1 Plan for students to ask questions | “I’m going to let them create their own investigation from their wonderings instead of the focus question from the science kit.” | They’re the questions makers. They decide. I want to put as much in their hands as possible.” |
| 2.2 Plan for students to ask testable questions | <i>No data</i> | “If you changed the variable, what would the question be?” |
| 2.3 Plan for students to design investigation | “They’re required to create the procedure. I’m not telling them what they should be doing.” | “I need to be sure that they are excellent at doing experiments.” |
| 2.4 Plan for students to design fair test | <i>No data</i> | “So when we do investigations, we need to have some parameters [controlling of variables].” “I would question how do you know you are only testing one variable.” |
| 2.5 Plan for students to focus on collecting data | “Then, they do a data observation chart or a descriptive piece.” | “I’m planning for good observation. I’m planning for good recording.” “They’re on their own to collect data.” “They learned from the mistakes that they made before.” |
| Plan for Evidence | | |
| 3.1 Plan for students to support their claims with evidence from investigations | “It’s taking the data and putting it in and analyzing it and learning from it.” “You need to provide evidence and tell them where you got those ideas from.” | “I’m planning for good written explanations of what they see, that explaining, that being able to speak what they’re seeing.” “What is happening in this circuit? I need you to explain it.” |

Table 1. Cont.

| Initial Codes | Examples from Ann | Examples from Lee |
|---|--|--|
| Plan for Evidence | | |
| 3.2 Plan for evidence to be reported out, but not questioned or evaluated | “I remind them, you know, ‘claims and evidence.’ What you claim, you need evidence.” | “What did you find? What information do you want to share?” “I need them to represent it in all different ways.” “You collect the data, get the median, and show me the data in graph form.” |
| Plan for Argument | | |
| 4.1 Plan for students to construct arguments for their claims | “They know I’m going to ask, ‘Why do you think that?’” “They are getting to the point in their conversations, ‘I agree with ___ because and I disagree with ___ because.’” | <i>No data</i> |
| 4.2 Plan for students to evaluate evidence when making claims | “They were able, after working with each other’s information...to figure out what was accurate and what was not accurate.” | <i>No data</i> |
| 4.3 Plan for students to consider alternative arguments | “They have to listen so they can either add on or disagree or agree with it” “They know I’m going to ask...’What do you think of his thinking?’” | <i>No data</i> |

In addition, the researcher had weekly discussions with each participant about emerging themes, explaining her science planning to ascertain the accuracy of the proposed ideas [65,78]. For example, each teacher responded by prioritizing, revising and expanding upon the researcher’s suggestion of factors that appeared to inform her planning in response to the NECAP science test. For a final member check, the participants reviewed the themes, the supporting evidence and quotations for their respective cases and documented their responses and corrections.

For qualitative inquiry, the goal is not to control variables or generalize to the larger population [65]. Rather, it is “to describe what people do and say in local contexts” to generate theory ([790], p. 29). In this case, themes were developed from evidence applied to Zembal-Saul’s [38] *a priori* theory of the continuum of teaching science, as well as from emerging data. By representing the teachers’ planning through detailed descriptions and by considering the factors affecting each teacher’s decisions, readers can determine the applicability to their own settings [79].

4. Research Context

Social phenomena are context-specific, operating within complex inter-related forces [65]. Examination of the historical background, the expectations and the resources in their social settings was employed to enhance the understanding of the planning decisions made by each of the teachers. The research context was described for three areas: (a) background and the format of the NECAP science test, (b) professional development available from their districts or that which they sought out on their own and (c) the nature of the science kit used by each teacher.

4.1. Background and Format of the NECAP Science Assessment

Prior to the 2008 NCLB mandate for states to implement measures of students’ science learning, this US state did not assess the science achievement of its students. The two teachers, Ann and Lee, indicated that before the NECAP science testing, they taught science from a textbook, expecting students to read and answer questions. Both teachers lacked confidence in teaching science and considered literacy to be their strongest area.

With the implementation of science testing, the state released science test items each year to teachers to inform them of the nature of the assessment. The NECAP assesses four domains: physical science, earth space science, life science and scientific inquiry [80]. For the three content areas, the test primarily utilizes the multiple choice format with some short answer questions. However, to “measure the student’s ability to make connections, express ideas, and provide evidence of scientific thinking” ([80], p. 10), the test also includes an “inquiry task” requiring students to provide performance-based constructed responses in 13 construct areas (see Table 2). Students conduct an investigation in small teams and then complete questions individually using the collected data. For example, the 2011 sample inquiry task posed the question: “How does increasing soil particle size affect the amount of water soil holds?” ([81], p. 2). Student teams conduct three investigations collecting data to measure the amount of water held by soil with small, medium and large particle sizes. Next, students individually complete constructed-response items: make a bar graph comparing the amount of water held by the three kinds of soil; describe the pattern in the graph; and explain how increasing soil particle size affects the amount of water soil holds. Lastly, the test booklet provides data from a similar study and information on growing conditions for cacti and ferns. Students use the data to identify the soil best suited for growing cacti and explain their rationale.

Table 2. NECAP constructs for broad areas of inquiry.

| Broad area of inquiry | Constructs to assess students’ ability with science processes |
|---|--|
| Formulating questions and hypothesizing | <ol style="list-style-type: none"> 1. Analyze information from observations, research, or experimental data for the purpose of formulating a question, hypothesis or prediction. 2. Construct coherent argument in support of a question, hypothesis, prediction. 3. Make and describe observations in order to ask questions, hypothesize, make predictions. |
| Planning and critiquing of investigations | <ol style="list-style-type: none"> 4. Identify information. Evidence that needs to be collected in order to answer the question, hypothesis, prediction. 5. Develop an organized and logical approach to investigating the question, including controlling variables. 6. Provide reasoning for appropriateness of materials, tools, procedures and scale used in the investigation. |
| Conducting investigations | <ol style="list-style-type: none"> 7. Follow procedures for collecting and recording qualitative or quantitative data, using equipment or measurement devices accurately. 8. Use accepted methods for organizing, representing, manipulating data. 9. Collect sufficient data to study question, hypothesis or relationships. 10. Summarize results based on data. |
| Developing and evaluating explanations | <ol style="list-style-type: none"> 11. Analyze data, including determining if data are relevant, artifact, irrelevant or anomalous. 12. Use evidence to support and justify interpretations and conclusions or explain how the evidence refutes the hypothesis. 13. Communicate how scientific knowledge applies to explain results, propose further investigations or construct and analyze alternative explanations. [82] |

A review of the 2008–2011 inquiry task items released to teachers indicated three types of responses required of students: provide explanations, represent data or describe an investigational design [81, 83–85].

4.1.1. Provide Explanations

Of the 31 sample questions, 71% called for students to provide a written *explanation based on evidence* for predictions, investigational designs, data analysis, conclusions or application to new situations. Though one broad area of inquiry was devoted specifically to “developing and evaluating explanation,” each of the four areas of inquiry required this skill (see Table 3). Words indicating a required explanation have been italicized.

Table 3. NECAP inquiry task questions for the four broad areas of inquiry requiring explanations.

| Broad area of inquiry | Example of NECAP released items questions for the Inquiry Task requiring students to provide an explanation |
|---|--|
| Formulating questions and hypothesizing | “Based on what you learned in your investigation, predict which food(s) a bird with this type of beak would eat and <i>explain why</i> ” ([83], p. 58). |
| Planning and critiquing of investigations | “ <i>Explain why</i> the models used in these investigations can be used to study how wind changes sand dunes. Use your data and observations <i>to support your answer</i> .” ([85], p. 43) |
| Conducting investigations | “Describe the pattern in your graph. <i>Explain how</i> increasing soil particle size affects the amount of water soil holds.” ([81], p. 32) |
| Developing and evaluating explanations | “Use what you learned in your investigation and what you know about what birds eat to <i>explain how</i> the shape of a bird’s beak affects its survival.” ([83], p. 54) “ <i>Explain what</i> people could do to keep the dunes at the beach in place. Use your data to <i>explain why</i> this would work.” ([85], p. 47) |

4.1.2. Represent Data

The second most common response required of students (19%) was to collect and represent data, found within the “conducting investigations” broad area of inquiry. For each of the four testing years, students were asked to create a bar graph of their data. An example from the 2009 NECAP released items included, “Make a bar graph that shows the data you collected. Graph the median numbers of pennies it took to move the box with no added weight, the box with the small weight, and the box with the large weight” ([84], p. 16).

4.1.3. Describe an Investigational Design

The fewest questions (10%) involved the student response of describing an investigational design. In the 2008 NECAP released items, students were asked to design their own investigation from a hypothetical scenario: fourth grade students were to determine if the shape of a bird’s beak was related to what it eats. Two tasks were posed: “Which beak will pick up the most different kinds of food? a. Write a plan...students can follow to help them answer their question. b. Identify one thing in your plan that will stay the same in the investigation” ([86], p. 2).

In summary, the NECAP inquiry task measured students' ability to conduct or design their own investigations, collect and represent their data, explain their rationale for experimental design decisions and justify their claims based on the collected evidence. Based on Zembal-Saul's [38] continuum for teaching science, the NECAP test assesses students' ability with investigation-based science, yet it emphasizes evidence-based explanation.

4.2. Professional Development

In 2008, Ann adopted the reform-based approach to teaching science when her district instituted a new science curriculum. Each elementary teacher received professional development and science consultant mentorship for four years to train them for the reform-based science kits. Additionally, Ann sought out professional development available through her district in Accountable Talk[®], a discourse approach for evidenced-based claims that she integrated into her science planning [87].

In Lee's district, science instruction was not an instructional priority. Though science kits were available to teachers, replenishment of supplies was intermittent, and no professional development or human supports were provided in science. In 2009 and 2010, Lee sought out professional development offered by a university based on the *Fundamentals of Inquiry* training [88] and received mentorship from the science education professor, Dr. Chen (pseudonym). This training focused on student ownership in raising questions, planning investigations, observing and interpreting data and hypothesizing based on evidence.

4.3. Science Kits

Ann used a Full Option Science System (FOSS) science kit, *Magnetism and Electricity*, over a ten week unit [89]. The FOSS science kit provided equipment, a teacher's guide, student worksheets, assessments, supplementary science stories, as well as a website with teacher resources, including content information, vocabulary and videotaped lessons of each investigation. In addition, the teacher's guide provided both teacher guided instructions to investigations or an open-ended approach, whereby students determined their own set-up to investigate the focus question.

For her six-week unit on electricity, Lee had a Science and Technology for Children (STC) kit on *Electric Circuits* from her district, providing equipment, a teacher's guide, student worksheets and assessments [90]. The teacher's guide provided instructions that teachers could follow to guide students in conducting investigations.

5. Results

The results of how teachers planned in response to the NECAP science assessment are presented through the lens of Zembal-Saul's [38] *continuum of teaching science as argument* to evaluate the level to which the teachers made instructional decisions aligned with reform science emphasizing evidence-based explanation and argumentation for student's critical thinking development. First, the constructs from the NECAP inquiry task released items are compared for alignment with NRC [9] standards-based scientific practices and the teachers' corresponding planning decisions. Next, the findings are reported for three major themes that explain the similarities and/or differences of Ann and

Lee's planning decisions for reform-based science in response to the state science assessment integrated with the training and resources they had available.

To gain an overview of the quality of the state science assessment and its impact on the teachers' planning, the constructs from the NECAP Inquiry Task released items were evaluated for alignment with NRC [9] standards-based scientific practices and compared with the teachers' instructional decisions. Table 4 depicts the following data for comparison purposes:

- Standards-based scientific practices for K-12 science education [9].
- The 13 Inquiry Task constructs on the NECAP science assessment.
- The instances of each construct provided to teachers as an Inquiry task released item by the state for each of the tested years from 2008–2011.
- Examples of quotations from each teacher as evidence of instructional decisions.

From this compilation of data, the evidence indicated that the 13 constructs for the inquiry task on the NECAP science assessment aligned with a standards-based scientific practice proposed by NRC [9]. However, there were no NECAP items released to teachers by the state for two scientific practices: modeling and argumentation. For the practice of “developing and using models,” there was no corresponding NECAP inquiry task construct, though assessment of students' understanding of the use of models was embedded in other released items (see Table 4). For the NRC practice of “engaging in argument from evidence,” no construct addressed argumentation involving evaluation of evidence or critique of claims from investigation data. Construct 2 stated “construct coherent argument in support of a question, hypothesis, prediction” *before* an investigation; yet, there were no corresponding released items for any of the years, 2008–2011, suggesting it was un-assessed. Construct 12 assessed students' ability to provide evidence-based *explanations*, not *argumentation*: “use evidence to support and justify interpretations and conclusions or explain how the evidence refutes the hypothesis.” This analysis suggested that the NECAP test assessed students' scientific practices consistent with the investigation and evidence-based levels of Zembal-Saul's [38] continuum. The evidence indicated that both teachers planned for their students to develop these investigation and explanation practices tested in the NECAP. However, from her training, Ann was conscious of and planned for building students' capacity to engage in argument from evidence resulting in her teaching science, also at the argument-based level. In contrast, Lee's plans did not include the scientific practice of argumentation, consistent with the absence of this practice on the NECAP released items.

Table 4. Alignment of science standards practices, NECAP inquiry task released items and teacher’s instructional decisions.

| Scientific Practices [9] | NECAP Standardized Science Test— Inquiry Task Constructs 1–13 | Construct Instances- Annual Inquiry Task Released Items | Ann—Sample Quote(s) of Instructional Decisions | Lee—Sample Quote(s) of Instructional Decisions |
|---|--|---|--|--|
| <i>Asking questions and defining problems</i> | 1. Analyze information from observations, research, or experimental data for the purpose of formulating a question, hypothesis or prediction. 3. Make and describe observations in order to ask questions, hypothesize, make predictions related to topic. | 1. (3)—2008, 2009, 2011 3. <i>None on Grade 4 Released Items—Years 2008–2011</i> | “I’m going to let them create their own investigation from their wonderings instead of the actual focus question from the science kit.” | “They’ve been stifled. I’ll get them knowing it’s comfortable and its OK...to question.” “They’re the question makers. They decide. I put as much in their hands as possible.” |
| <i>Developing and using models</i> | <i>No construct for this practice on the NECAP.</i> | <i>None on Grade 4 Released Items—Years 2008–2011</i> | “We defined what a series circuit is and they drew a picture with no materials about what they thought it would look like and what was going to happen.” | “Draw a series circuit that would be able to test if the flow goes through an object to close the circuit.” |
| <i>Planning and carrying out investigations</i> | 4. Identify information/ evidence that needs to be collected in order to answer the question, hypothesis, prediction. 5. Develop an organized and logical approach to investigating the question, including controlling variables. 6. Provide reasoning for appropriateness of materials, tools, procedures and scale used in the investigation. 7. Follow procedures for collecting and recording qualitative or quantitative data, using equipment or measurement devices accurately. 9. Collect sufficient data to study question, hypothesis or relationships. | 4. (1)—2010 5. (2)—2008, 2011 6. (2)—2009, 2010 7. (2)—2008, 2009 9. (1)—2011 | “I give them the focus question and they figure out how they want to solve it.” “They’re required to create the procedure. I’m not telling them what they should be doing. Then, they do a data observation chart or a descriptive piece.” “My plan is they’re going to set up [the series circuit] designed by every person individually in the small groups...and they’re going to be able to talk it through and see if its right...what worked, what didn’t work and why.” | “So when we do investigations, we need to have some parameters [controlling of variables].” “They’re on their own to collect data.” “They learned from the mistakes [of data collection] that they made before.” “Think with your partner, what do you want to try, what do you want to test or experiment with?” |

Table 4. Cont.

| Scientific Practices [9] | NECAP Standardized Science Test— Inquiry Task Constructs 1–13 | Construct Instances- Annual Inquiry Task Released Items | Ann—Sample Quote(s) of Instructional Decisions | Lee—Sample Quote(s) of Instructional Decisions |
|--|--|---|---|--|
| <i>Analyzing and interpreting data</i> | 11. Analyze data, including determining if data are relevant, artifact, irrelevant or anomalous. | 11. (2)—2008, 2011 | <p>“My focus is to slow down the amount of inquiry so we really delve into what is the data, what does it really mean.” “You did it this way and they did it this way. What is the difference in your data? Why do you think that? What are those things that you’re not realizing?”</p> <p>“They have to analyze it [the data], make decisions from it.”</p> | <p>““What looks different about this teams’?”</p> <p>“Let them notice there’s a difference and figure it out.”</p> |
| <i>Using mathematics and computational thinking</i> | 8. Use accepted methods for organizing, representing and manipulating data. | 8. (4)—2008, 2009, 2010, 2011 | <p>“They have to make representations of it [the data], make graphs....It’s taking the data and putting it in and analyzing it and learning from it.”</p> | <p>“I needed them to represent it [the data] in all different ways.” “You collect the data, get the median, and show me the data in graph form.” “They have to represent data accurately....really labeled well....and attention to detail.”</p> |
| <i>Constructing explanations and designing solutions</i> | <p>10. Summarize results based on data.</p> <p>12. Use evidence to support and justify interpretations and conclusions or explain how the evidence refutes the hypothesis.</p> | <p>10. (3)—2009, 2010, 2011</p> <p>12. (6)—2008, 2008, 2009, 2009, 2010, 2011</p> | <p>“You need to provide evidence and tell them where you got those ideas from.”</p> <p>“They know I’m going to ask, ‘Why do you think that?’” “What do you think of his thinking?”</p> | <p>“I’m planning for good written explanations of what they see, that explaining, that being able to speak what they’re seeing.” “What are your thoughts? What did you find? What information do you want to share?”</p> |

Table 4. *Cont.*

| Scientific Practices [9] | NECAP Standardized Science Test— Inquiry Task Constructs 1–13 | Construct Instances- Annual Inquiry Task Released Items | Ann—Sample Quote(s) of Instructional Decisions | Lee—Sample Quote(s) of Instructional Decisions |
|---|--|--|---|---|
| <i>Engaging in argument from evidence</i> | 2. Construct coherent argument in support of a question, hypothesis, prediction. | <i>None on Grade 4 Released Items—Years 2008–2011</i> | “My goal is student to student accountable talk...they have to listen so they can either add on or disagree or agree with it—‘We saw that too, and this is what we did next.’” | |
| <i>Obtaining, evaluating, and communicating information</i> | 13. Communicate how scientific knowledge applies to explain results, propose further investigations or construct and analyze alternative explanations. | 13. (3)—2009, 2010, 2010 | “We’re trying to get them...to apply what they’re learning.” “After working with each other’s information, they were able to figure out what was accurate and what was not accurate.” “They are getting to the point in their conversations, ‘I agree with ___ because and I disagree with ___ because,’ but they need to put that into their writing too.” | “I want them to solve problems and I want them to apply their knowledge.” “I require more explanation not only verbally but through writing...sometimes it’s graphs, sometimes, it’s pictures.” |

In reviewing the data collected from Ann and Lee regarding their planning decisions, three major themes emerged to explain their planning decisions for reform-based science in response to the NECAP science assessment and to depict the similarities and differences in their planning.

1. *Science planning informed by the state science assessment.* Despite pressure in high stakes subject areas, both teachers were committed to providing reform-based science. Though there were similarities in their approach to planning, the results indicated differences in their goals and focus areas informed by their respective training and resources.
2. *Student ownership of investigations.* Both teachers promoted student ownership of the investigations, informed by the tasks and constructed-response format students were expected to complete on the test reflecting the investigation-based level [38]. However, the emphases on promoting students' development of testable questions, controlled variables and fair tests varied between the teachers, which correlated with the nature of the professional development each received in investigation-based science.
3. *Student scientific discourse as precursor to writing.* Both teachers promoted students' scientific discourse as a means to build their capacity to generate and write evidence-based conclusions on the test reflecting the evidence-based level of teaching science [38]. However, only Ann provided the argument-based level of teaching science, informed by the training she received in Accountable Talk[®].

Each theme is presented for the two cases using "particular description", including events and the participant's own words as evidence to warrant the assertions ([77], p. 149).

5.1. Science Planning Informed by the State Science Assessment

In their urban settings, though both teachers experienced pressure to increase their students' achievement in the high-stakes subjects of reading and mathematics for the district to meet AYP, they provided reform-based science with the inception of the NECAP science testing. For example, Lee noted that "science doesn't count for accountability", but "the NECAP is directly in the back of my mind all the time." Ann explained, "I think the science test benefits the kids even on the reading test in the fall because it's again the writing and the analyzing of literature. So it's all interconnected." The teachers also recognized how the assessment was leading to improved science learning for their students. Ann noted, "I think it is pushing my science teaching to the next level. Maybe I wouldn't be so dedicated to challenging them if it wasn't being assessed." Lee expressed, "Our inner cities are penalized to not give the beautiful science that we can provide for them because our reading scores are low? The affluent communities get so much better science. That's a shame." She was committed to providing reform-based science to help her students develop as "free thinkers." The teachers decided upon focus areas for the science unit, in part, based on their understanding of the NECAP science test items released by the state.

An analysis of the data indicated that both teachers planned their science lessons to prepare their students for the inquiry task format rather than the multiple choice content questions. The teachers felt students would acquire content knowledge through their investigations. They focused on the inquiry

task as the more difficult aspect of the NECAP state science assessment when planning science lessons. However, their selection of instructional goals from the inquiry task differed.

Ann concentrated on building students' capacity to write evidence-based conclusions from their investigations:

"They're great at doing the Inquiry Task in their small group.... So we're not so focused on the content knowledge....We need to work on their data and their analysis and their written output of what they've really learned. It's analyzing the data and writing about it. That's where we're falling apart."

Ann felt that conclusion writing was so difficult for students to master that she wanted to start early in the year. The science consultant in her school also had stressed to her the importance of building students' capacity with writing in science as a means for students to improve their scores on the NECAP inquiry task. Ann would introduce experimental design and data representation in this unit and continue to incorporate these skills in her mathematics instruction and during later science units.

From Lee's understanding of the NECAP science test and her training in inquiry-based science with Dr. Chen, she felt she needed to prepare her students to be "excellent at experiments, excellent data collectors, excellent data showers, whether it be in a graph or words and then, I could put any experiment in front of them and they start to extrapolate great information from it, and that's the content." Lee decided to focus first on data collection and representation, since elementary students statewide "fell short on putting together a graph properly." Yet, she also viewed the inquiry task as focused on "application" of information from the investigation to new situations.

"I realize that the NECAP is application, whereas with the textbook, I saw content....I need them to apply their knowledge to real life. That's what the NECAP does. So I'm always telling them they're going to ask you, "Now you have this knowledge, apply it." I think I was missing the application piece before."

Lee planned challenges for students to apply the evidence they found from their investigations and explain their thinking.

Thus, each teacher was aware of the nature of the inquiry task questions and set goals for the science unit through the lens of the particular training and/or resources each had available. Ann focused on written evidence-based explanations, as well as data collection/analysis. Lee targeted student's data collection/analysis and application of their findings to real life. Despite differences in their goals, the science planning for both teachers was consistent with the investigation- and evidence-based levels of Zembal-Saul's continuum [38].

5.2. Student Ownership of Investigations

Both teachers recognized that students were required to conduct an investigation for the NECAP science inquiry task and planned for students to take ownership of the investigations. Three sub-themes emerged from the data analysis indicating how the teachers planned to increase student responsibility in investigation-based practices.

- Promotion of student capacity to investigate scientific questions

- Adoption of an “open” investigation approach to the science kit
- Development of students’ capacity in writing for investigations

The teachers focused on students’ development in scientific practices of asking questions and defining problems, planning and carrying out investigations, analyzing and interpreting data and using mathematics and computational thinking. However, there were differences in the emphasis that each teacher placed on specific aspects of the practices.

5.2.1. Promotion of Student Capacity to Investigate Scientific Questions

Ann designed her lessons to develop students’ independent role in asking questions. As the unit progressed, she “let them create their own inquiry task from...their wonderings” and design and conduct an investigation on their own. Ann’s planning was based on the belief that students working in small groups could develop their own investigational procedure to answer a focus question, represent their data, monitor team member’s graphing of results and support each other’s learning. For example, during an observation of a lesson, student teams generated their own questions and wanted to follow-up on an investigation during the next lesson. As a result, Ann made the decision to let them “use their wonderings to create a focus question for tomorrow on their own.” She found that her students “really flourished having that responsibility.” Ann’s emphasis was on providing students with a safe environment to share their “wonderings” and empower students to develop their own investigational procedures. However, she did not plan specifically for students to evaluate whether their questions were testable or whether they were designing a fair test (see Table 1).

Lee’s planning also centered on increasing her students’ ownership of investigations. In contrast with Ann, she made instructional decisions to “grow their skills” in developing testable questions, designing controlled investigations and determining data collection and representational approaches (see Table 1). For example, knowing that the test scores “were really low with data representation,” Lee planned purposefully for her students to be “on their own to collect data” for a first investigation requiring extensive data recording and mathematical calculations. In class discussion, students decided which methods were “more efficient”; they “learned from their mistakes” by creating their own organization systems and “through questioning and showing and sharing.” Lee had acquired pedagogical knowledge in promoting these practices through her inquiry-based science training.

The planning choices made by both teachers to support students’ science reform learning were informed by three of the four broad areas of inquiry on the NECAP inquiry task of *formulating questions, planning investigations and conducting investigations*, consistent with the investigation-based level of Zembal-Saul’s continuum [38].

5.2.2. Adoption of an “Open” Investigation Approach to the Science Kit

To assist their students in developing independence with scientific practices needed for the NECAP inquiry task, both teachers adopted a more open, student-centered approach to the science kit.

The evidence suggested that the FOSS science kit was a resource for Ann’s planning. The teacher’s guide offered two pedagogical options: a teacher-guided approach to investigations or an open-ended approach, whereby students designed procedures to investigate focus questions [89]. As the unit

progressed, Ann chose the open-ended approach, allowing students to “talk it through” and “figure out how they were going to do it.” The science consultant was a support for Ann in planning for student-centered implementation of the science kit lessons. This approach built a class culture that student thinking was valued, students planned their investigations and student “wonderings” could generate future investigations.

The teacher’s guide was less of a resource for Lee, since she felt “the teacher’s edition stifles me. When I hang my hat on that, I don’t allow anybody to move anywhere but within the parameters of the actual lesson.” She had acquired training in reshaping science kit lessons from her mentor, Dr. Chen. Thus, Lee chose lessons from the STC kit on electric circuits and redesigned them for more student freedom in determining the question to be investigated and/or the procedure that they would use to solve a problem. She found students acquired deeper learning from this approach, particularly for the 50% of her students who had learning differences or who struggled with reading.

From knowledge that the NECAP inquiry task focused on formulating questions and planning/conducting investigations, both teachers made decisions to prepare students to be independent with these practices by adopting an open, student-centered approach.

5.2.3. Development of Students’ Capacity in Writing for Investigations

Ann and Lee participated in different trainings for reform-based science. Ann’s training focused on how to implement the FOSS science kit [89]; whereas Lee participated in an independent training on inquiry-based science. These different opportunities impacted how each teacher decided to prepare students for written tasks on the science test.

Given that Ann’s primary goal for the first unit was to increase student capacity with evidence-based conclusion writing, Ann chose to use the science kits’ student worksheets with challenge questions, because she found she was “getting more good solid writing out of them.” She expressed, “I think they’re so worth it...the groundwork...is really paying off, the modeling and the worksheets” to prepare students for writing on the NECAP inquiry task.

In contrast, Lee chose to promote students’ generation of their own written formats to represent data for analysis. Through her inquiry training with Dr. Chen, Lee learned about “different levels of teacher intervention” and how to promote students’ construction of their own data collection/representation formats, whether it be a T-chart, Venn diagram or a written explanation. As a result of this training, Lee decided to “put more on my students and not so much structure of a worksheet from me” for her students to make sense of the data.

Thus, when planning for student writing to analyze data, Ann concentrated on developing explanations using science kit worksheets, while Lee focused on student ownership of data representation, interpretation and explanation. Both approaches to writing required critical thinking; yet, these decisions were a product of each teacher’s view of the inquiry task in conjunction with the pedagogical knowledge they gained from their resources and training.

5.3. Student Scientific Discourse as a Precursor to Writing

Both teachers understood that to provide evidence-based explanations from investigation results, the students needed confidence in expressing themselves and language for scientific discourse. Three

sub-themes emerged, depicting the teachers planning to promote student's discourse as a means to increase their meaning making for the writing that would be required of them on the NECAP inquiry task.

- Development of students' scientific vocabulary
- Promotion of oral discussions and student writing for evidence-based explanations
- Creation of a climate for scientific discourse and argumentation

However, the teachers' background in scientific discourse varied. Ann had training in engaging students in evidence-based discourse and argumentation through Accountable Talk[®], whereas Lee's training did not stress this practice.

5.3.1. Development of Students' Scientific Vocabulary

Ann was aware her students had limited exposure to science vocabulary and experience in expressing scientific ideas orally, explaining that students entering first grade in urban schools typically having one-third the number of words of students in more affluent districts. Thus, Ann routinely prompted students: "I will be giving you some vocabulary just as on the NECAP. I want you to use these words during your inquiry task and then in your writing." From this decision, she found students increased their use of science language. For example, during an observation of a lesson in which the students were challenged to make an electromagnet from wires, a rivet and a battery to pick up metal washers, one student questioned how electricity could flow through the rivet: "I'm confused. If there's an insulator wrapped around the wire, how can it go through?" Another student responded by explaining it had become "a temporary magnet." Ann explained that this student typically "has a hard time expressing himself," yet he was able to recall the understandings and language used in a previous lesson. Ann noted that her decision to emphasize vocabulary acquisition had "paid off because he carried over" his learning to make a connection between their study of magnets and electricity.

Lee was aware that her students "get a higher score if you apply the proper vocabulary to your answer" on the NECAP test. She believed that if she "scaffolded the vocabulary well, it would become innate" for her students. During an observation of a lesson in which students made a flashlight to demonstrate their understanding of electric circuits and switches, Lee prompted a boy to use science language in his explanations. She noted that she consciously plans for prompting students to use scientific language in their speech as preparation for their using scientific vocabulary in their writing on the inquiry task.

5.3.2. Promotion of Oral Discussions and Student Writing for Evidence-Based Explanations

Both teachers valued oral discussions as a means for their students to make meaning from the data they collected in their investigations. They felt that students could express themselves in writing more easily after first processing their thinking orally.

Ann believed that an essential practice of science was developing evidence-based explanations for investigation results. She would remind students, "You need to provide evidence and tell them where you got those ideas." She would prompt students to use language, such as, "I claim this because...and I know this because..." Likewise, for their written conclusions, Ann reminded students they needed

evidence for their claims: “Why do you think that, why?” Ann’s questioning resulted in a pattern of talking and writing that students expected. She established a procedure for students first to explain and debate their findings with their group and then communicate their conclusions in writing.

Lee planned for her students to share the understanding they constructed from their small group investigations. By student teams comparing the results from their self-designed investigations, the discussion promoted students’ meaning-making and construction of understanding about circuitry. Lee explained, “I’m planning for good written explanations of what they see, that explaining, that being able to speak what they’re seeing.” Since Lee knew her students were “intimidated” by writing, she encouraged students’ oral communication before they notated their thinking on paper.

The teachers’ decisions to concentrate first on oral communication laid the groundwork for students to express their thoughts in writing in preparation for the constructed responses on the inquiry task. Both teachers promoted students use of evidence to support and justify conclusions, consistent with Zembal-Saul’s evidence-based science teaching [38].

5.3.3. Climate for Scientific Discourse and Argumentation

Based on the evidence, only Ann achieved Zembal-Saul’s argument-based science teaching [38], involving the evaluativist level of critical thinking (see Table 1). Lee had not acquired knowledge of the scientific practice of engaging in argument from evidence, which could support students in evaluating and communicating findings. Her training did not focus on argumentation to critique explanations. Thus, Lee’s decisions were limited by the resources and training she had at her disposal.

Ann’s training in Accountable Talk[®] informed her planning for scientific discourse. She modeled “talk moves”, such as: explicating reasoning, restating someone else’s reasoning, applying own reasoning to someone else’s reasoning, challenging someone else’s reasoning or providing a counter example [87]. Ann explained, “I view it [Accountable Talk[®]] as a powerful discussion technique” that students internalize in their conversations: “They’re agreeing with each other. They’re building off each other’s ideas or they’re disagreeing and telling them why.” Ann prompted students to listen to each other, because she believed that students could resolve misconceptions by physically demonstrating their thinking and considering alternatives with their group members. She discovered that by giving students more “time” for discussion, the thinking within a group would emerge, “Well, we know we can do it this way”, and another group would add, “Well, we know you can also do it this way.” Students from different groups would discuss their results, as observed in one lesson, and develop new questions to investigate, because they realized they needed to gather additional data and compare the results from the new investigations. Through scientific discourse, students began to acquire scientific practices of constructing explanations, critiquing claims and considering alternative ideas—practices aligned with the standards-based intent of explanation and argumentation. Though the NECAP did not include test items in argumentation, Ann acquired the knowledge for these practices from her training in Accountable Talk[®] moving her to the argument-based level of Zembal-Saul’s continuum [38].

In summary, Ann and Lee’s planning to develop students’ reform-based science practices was grounded in an awareness of the actions required of students on the NECAP inquiry task. Yet, their planning decisions were made within the lens of their professional development training and available

resources. These urban teachers were acquiring skill in promoting their students' critical thinking and problem-solving through scientific practices and discourse, to the extent they were aware, with fidelity to the science standards.

6. Discussion and Conclusion

The purpose of this study was to examine the instructional planning by two elementary teachers in urban schools for reform-based science in response to a state science assessment within the context of their available training and resources. Qualitative methods were used to conduct an in-depth analysis in order to gain understanding of how these teachers provided equitable pedagogy for their students' critical thinking development through reform-based science given the pressured accountability climate in high-poverty urban schools. Since the inception of the 2008 mandated state science testing, these two veteran teachers, who previously taught science from a textbook, were transforming their science teaching practice to be in alignment with the science standards intent for students' scientific literacy.

The findings suggested that their planning and goal-setting for reform-based instruction was informed, in part, by the inquiry task on the high-quality NECAP state science test [14], as well as by their understanding of science teaching from their professional development training and science kit resources. The NECAP inquiry task required teams of students to conduct a scientific investigation followed by individual application, interpretation and explanation based on evidence [82]; practices consistent with the intent of standards-based learning for collaboration, problem solving, analysis and critical thinking [19]. The teachers indicated that the expectations for the inquiry task, rather than the multiple choice content questions, guided their science instructional planning, because they considered this aspect of the science test to be most difficult for their students. From their understanding of the inquiry task question types, they planned for students' development of scientific practices not only in asking questions, planning and carrying out investigations, analyzing and interpreting data and using mathematics and computational thinking, but also in constructing and communicating evidence-based explanations. Though researchers have found that elementary teachers tend to concentrate on the *procedures* of science experimentation [9,23,41], the primary type of response required of students on the inquiry task to *provide explanations* influenced these teachers to plan for students' higher-order thinking and meaning making from the investigations.

Thus, this researcher argues that the form of assessment in the high-quality inquiry task made a difference in guiding the type of science instruction the two teachers provided to their students, even in the face of district pressure to allocate instructional time to high-stakes subjects of reading and math. The evidence from this study supports literature in the field of assessment that elementary teachers use state test results to make decisions about their instruction [61] and that some standardized test formats can be associated with teachers' focus on developing students' critical thinking skills [53]. For these two teachers, the inquiry task format served to guide them to plan for reform-based scientific practices consistent with Zemba-Saul's investigation-based and evidence-based science teaching [38], rather than rote learning of content. The science assessment presented an expectation for student performance aligned with the science standards from which the teachers planned backwards [91].

However, in order to operationalize the goals they determined from the inquiry task, both teachers relied on their professional development training and resources. The second argument posited from

this research is that the nature of each teacher's professional development and resources provided a lens through which they determined goals in preparing students for the state science assessment. In terms of Zembal-Saul's *continuum for teaching science as argument* [38], each teacher's advancement of her teaching practice depended on the understanding she had acquired of reform-based science. From Lee's training in inquiry science, of which one aspect was data collection/representation, she was able to address her awareness of her students' weakness in graphing on the science assessment by targeting a goal to build her students' ownership of data representation. Ann and Lee's training in investigation-based science informed them of focus areas for science instruction and gave them confidence to revise science kit instructions or select more open-approaches for students to design investigations, grapple with data and share evidence-based explanations. The teachers' training had enhanced their ability to read into a teacher-guided approach recommended from the science kit and envision a student-centered method. Ann and Lee's decisions to adjust their instruction over time from teacher-guided to more open, student-centered investigations is consistent with research suggesting that effective teachers in urban schools incrementally build up their student's ability to make connections for themselves [44,92]. Thus, the training that both teachers had received in conjunction with their science kit curriculum materials supported their self-efficacy in providing learning experiences for students that reflected both investigation-based and evidence-based science teaching, according to Zembal-Saul's continuum [38].

However, the study's evidence also supported a third argument that for teachers to enact the full range of NRC standards-based scientific practices in the classroom [9], they needed explicit professional development in each area. For example, though both teachers recognized that students needed to conduct investigations, display and interpret their data and develop a question and experimental design for an application question on the NECAP inquiry task, only Lee had training in promoting students' generation of testable questions and controlled fair tests and planned for these practices in her science classroom. Without that explicit training, Ann's planning for student ownership of investigations was more generalized, providing them with the opportunity to address their wonderings and their own investigational procedures without considering the rigor of those practices.

Conversely, only Ann was able to incorporate argumentation into her pedagogical plans, specifically from her professional development training in Accountable Talk[®], which promotes students' evidence-based discourse [87]. Given that the test did not assess this practice, the NECAP assessment was not responsible for her adoption of this pedagogical approach. Though Accountable Talk[®] is a general teaching strategy not specific to science, used to engage students in evidence-based talk and critique, Ann's training increased her awareness of how it could prepare students for the inquiry task and provided a means for how she could incorporate it into her science planning to support students' meaning making from their investigations. She was able to provide argument-based science teaching for her students [38], consistent with research in cognition, indicating that teachers can nurture young children's capacity to explain, reason and debate their claims [20,93-95] in the urban elementary science classroom through scientific discourse [43] and class participation structures [46]. Furthermore, evidence that argumentation was incorporated in Ann's discourse practices and not Lee's is in accordance with Osborne *et al.*'s [34] assertion that teachers can change their classroom discourse when they are supported explicitly with materials and pedagogical strategies to teach students the skill of reasoning through argumentation.

Though research has indicated that students' learning together using group investigation methods can result in significantly higher achievement than competitive or individualistic learning [96], classroom learning approaches used by both teachers, Mercer, Dawes, Wegerif and Sams have reported that children cannot be expected to engage in reasoned discourse effectively with their peers if they have not been helped to learn "how to talk together and for what purpose" ([97], p. 361). Mercer *et al.* found that as elementary students increased their use of language in practices, such as describing observations clearly, justifying their views, seeking reasons for claims and critically examining competing explanations, they learned and acquired deeper conceptual understanding in science [97]. Thus, Ann's training in specific "talk moves" [87] provided her not only with awareness of the language for scientific discourse, but also the impetus to model and expect this form of conversation from herself and her students. Though Lee employed small group and class discussions for students to present evidence from investigations, she had not learned discourse strategies for students to evaluate their evidence or critique claims that would support her science teaching at Zembal-Saul's argument-based level [38] and her students' further acquisition of critical thinking skills.

Though the conclusions drawn from this study are context-specific for the limited sample of two teachers [71], the implications from this research suggest that the form of an assessment could make a difference in a teacher's instructional choices and alignment of student learning with the standards, even in highly pressured urban schools. The current science standards' emphasize explanation and argumentation [9]; however, the results of this study point out the absence of argumentation questions targeted to evaluation of evidence and claims, in spite of the high-quality inquiry nature of the NECAP science assessment [14]. This finding conveys the need for assessment designs to more fully match the intentions of the standards in assessing students' higher-order thinking.

Furthermore, the results of this research suggest that attention be given to the nature of teacher's professional development and resources available for reform-based science and whether these teacher supports are fully aligned to the science standards. The evidence from this study of two teachers demonstrated how the teachers were able to plan for science practices only to the extent they were aware of these reforms. Particularly in urban schools, where teachers are expected to adopt a plethora of educational initiatives in efforts to increase students' assessment scores in reading and math, it is essential that the supports they do receive for science education are consistent with the NRC practices for all students to acquire scientific literacy [9,10].

In light of current reforms in science education, further studies that include rich stories from other teachers in urban schools with different state science assessment formats, training and resources can increase the knowledge base of factors that support, as well as impede, reform-based science teaching/learning for critical thinking. Examination of teacher's science planning decisions if the inquiry task were eliminated would shed more light on the impact of assessment format on critical thinking pedagogy. Furthermore, this research suggests the need to investigate whether urban teachers' implementation of reform-based science would increase with more targeted professional development in argumentation. Finally, this study focused on science; however, with educational reforms both in the US and internationally emphasizing the development of higher-order thinking skills for the twenty-first century, research is needed to investigate other content areas for both the alignment of assessments with their respective standards and the nature of teachers' professional development/resources,

particularly for reading and math given their high-stakes accountability status, and how these factors impact teachers' planning decisions for students to develop skills in problem-solving and critique.

In conclusion, this study contributed to the existing bodies of literature calling for coherent standards, assessment and instruction [15,53] and students' equitable access to critical thinking development [14,16,48,54,55]. Delpit, a researcher in the field of equitable education, highlighted the importance of empowering students who are marginalized in society by "teaching the linguistic aspects of the culture of power" and higher-level thinking skills to increase their future success in society ([29], p. 29). The present need in society for citizens to be knowledgeable in science for an increasingly technical and scientific workforce, as well as to evaluate national and global issues underscores the role of education in promoting the scientific literacy of all students [1,20]. This study took a step in describing the factors influencing decisions made by two urban teachers toward this goal. In this accountability climate, further research is needed on the types of assessments and training that support teachers' planning for reform-based science to promote their students' equitable learning for critical thinking.

Conflict of Interest

The author declares no conflict of interest.

References

1. National Research Council. *Taking Science to School: Learning and Teaching Science in Grades K-8*; The National Academies Press: Washington, DC, USA, 2007.
2. National Research Council. *National Science Education Standards*; The National Academies Press: Washington, DC, USA, 1996.
3. U.S. Department of Education (USDE). *Public Law 107-110, the No Child Left Behind Act of 2001*, 8 January 2002. Available online: <http://www.ed.gov/policy/elsec/leg/esea02/pg2.html#sec1111> (accessed on 4 October 2008).
4. National Science Foundation. *NST at a Glance*. Available online: <http://www.nsf.gov/about/glance.jsp> (accessed on 10 June 2013).
5. Lazerson, M. *American Education in the Twentieth Century*; Teachers College Press: New York, NY, USA, 1987.
6. Atkin, J.M.; Karplus, R. Discovery of invention? *The Sc. Teach.* **1962**, *29*, 45–51.
7. BSCS. *Science for Life and Living*; Kendall/Hunt Publishing: Dubuque, IA, USA, 1988.
8. National Research Council. *Inquiry and the National Science Education Standards: A Guide for Teaching and Learning*; The National Academies Press: Washington, DC, USA, 2000.
9. National Research Council. *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*; The National Academies Press: Washington, DC, USA, 2012.
10. NGSS. *Next Generation Science Standards: For States, by States*; Achieve Inc. on behalf of the 26 states and partners: Washington, DC, USA, 2013. Available online at <http://nextgenscience.org/next-generation-science-standards> (accessed on 10 June 2013).
11. Diamond, J.B.; Spillane, J.P. High-stakes accountability in urban elementary schools: Challenging or reproducing inequality? *Teach. Coll. Rec.* **2004**, *106*, 1145–1176.

12. Spillane, J.P.; Diamond, J.B.; Walker, L.J.; Halverson, R.; Jita, L. Urban school leadership for elementary science instruction: Identifying and activating resources in an undervalued school subject. *J. Res. Sci. Teach.* **2001**, *38*, 918–940.
13. Ravitch, D. *The Death and Life of the Great American School System: How Testing and Choice are Undermining Education*, Basic Books: New York, NY, USA, 2010.
14. Darling-Hammond, L. Race, inequality and educational accountability: The irony of ‘No Child Left Behind’. *Race. Ethn. Educ.* **2007**, *10*, 245–260.
15. Haertel, E.; Herman, J. *A Historical Perspective on Validity Arguments for Accountability Testing* (CSE Tech. Rep. No. 654); University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST): Los Angeles, CA, USA, 2005.
16. Hursh, D. Exacerbating inequality: The failed promise of the No Child Left Behind Act. *Race Ethn. Educ.* **2007**, *10*, 295–308.
17. Johnson, C. C. Effective science teaching, professional development and No Child Left Behind: Barriers, dilemmas, and reality. *J. Sci. Teach. Educ.* **2007**, *18*, 133–136.
18. Lee, O.; Krajcik, J. Large-scale interventions in science education for diverse student groups in varied educational settings. *J. Sci. Teach. Educ.* **2012**, *49*, 271–280.
19. Darling-Hammond, L. *Performance Counts: Assessment Systems that Support High-Quality Learning*; Council of Chief State School Officers: Washington, DC, USA, 2010.
20. Kuhn, D. A developmental model of critical thinking. *Educ. Res.* **1999**, *28*, 16–26, 46.
21. Kuhn, D.; Weinstock, M. What Is Epistemological Thinking and Why does It Matter? In *Personal Epistemology: The Psychology of Beliefs about Knowledge and Knowing*; Hofer, B., Pintrich, P., Eds.; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 2002; pp. 121–144.
22. Brown, A.L. Metacognition, Executive Control, Self-regulation, and Other More Mysterious Mechanisms. In *Metacognition, Motivation, and Understanding*; Weinert, F.E., Kluwe, R.H., Eds.; Lawrence Erlbaum: Hillsdale, NJ, USA, 1987; pp. 65–116.
23. Abd-El-Khalick, F., Boujaoude, S., Duschl, R., Lederman, N.G., Mamlok-Naaman, R., Hofstein, A., Niaz, M., Treagust, D., Tuan, H. Inquiry in science education: International perspective. *Sc. Educ.* **2004**, *88*, 397–419.
24. Bryan, L.A. Nestedness of beliefs: Examining a prospective elementary teacher’s belief system about science teaching and learning. *J. Res. Sci. Teach.* **2003**, *40*, 835–868.
25. King, K.; Shumow, L.; Lietz, S. Science education in an urban elementary school: Case studies of teacher beliefs and classroom practices. *Sci. Educ.* **2001**, *85*, 89–110.
26. Lee, O.; Hart, J.E.; Cuevas, P.; Enders, C. Professional development in inquiry-based science for elementary teachers of diverse student groups. *J. Res. Sci. Teach.* **2004**, *41*, 1021–1043.
27. Yerrick, R.K.; Hoving, T.J. One foot on the dock and one foot on the boat: Differences among preservice science teachers’ interpretations of field-based science methods in culturally diverse contexts. *Sci. Educ.* **2003**, *87*, 390–418.

28. Anyon, J. What Should Count as Educational Research: Notes toward a New Paradigm. In *Education Research in the Public Interest: Social Justice, Action, and Policy*; Ladson-Billings, G., Tate, W.F., Eds.; Teachers College Press: New York, NY, USA, 2006; pp. 17–26.
29. Delpit, L. *Other People's Children: Cultural Conflict in the Classroom*; The New Press: New York, NY, USA, 1995.
30. Duschl, R.A. *Restructuring Science Education: The Importance of Theories and Their Development*; Teachers College Press: New York, NY, USA, 1990.
31. McNeil, K.L.; Lizotte, D.J.; Krajcik, J.; Marx, R.W. Supporting students' construction of scientific explanations by fading scaffolding in instructional materials. *J. Learn. Sci.* **2006**, *15*, 153–191.
32. Sandoval, W.A.; Reiser, B.J. Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Sci. Educ.* **2004**, *88*, 345–372.
33. Driver, R.; Newton, P.; Osborne, J. Establishing the norms of scientific argumentation in classrooms. *Sci. Educ.* **2000**, *84*, 287–312.
34. Osborne, J.; Erduran, S.; Simon, S. Enhancing the quality of argumentation in school science. *J. Res. Sci. Teach.* **2004**, *41*, 994–1020.
35. American Association for the Advancement of Science. *Science for All Americans: Education for a Changing Future*; Oxford University Press: New York, NY, USA, 1990.
36. Barrow, L.H. A brief history of inquiry: From Dewey to standards. *J. Sci. Teach. Educ.* **2006**, *17*, 265–278.
37. Minner, D.D.; Levy, A.J.; Century, J. Inquiry-based science instruction—What is it and does it matter? Results from a research synthesis years 1984–2002. *J. Res. Sci. Teach.* **2010**, *47*, 474–496.
38. Zembal-Saul, C. Learning to teach elementary school science as argument. *Sci. Educ.* **2009**, *93*, 687–719.
39. Linn, M.C. Designing the knowledge integration environment. *Int. J. Sci. Educ.* **2000**, *22*, 781–796.
40. Thadani, V.; Cook, M.S.; Griffis, K.; Wise, J.A.; Blakey, A. The possibilities and limitations of curriculum-based science inquiry interventions for challenging the “pedagogy of poverty”. *Eq. Excell. Educ.* **2010**, *43*, 21–37.
41. Schibeci, R.A.; Hickey, R. Is it natural or processed? Elementary school teachers and conceptions about materials. *J. Res. Sci. Teach.* **2000**, *37*, 1154–1170.
42. Bartholomew, H.; Osborne, J.; Ratcliffe, M. Teaching students “ideas-about-science”: Five dimensions of effective practice. *Sci. Educ.* **2004**, *88*, 655–682.
43. Varelas, M.; Pappas, C.C.; Kane, J.M.; Arsenault, A. Urban primary-grade children think and talk science: Curricular and instructional practices that nurture participation and argumentation. *Sci. Educ.* **2008**, *92*, 65–95.
44. Flynn, N. What do effective teachers of literacy do? Subject knowledge and pedagogical choices for literacy. *Literacy* **2007**, *41*, 137–146.
45. Amaral, O.M.; Garrison, L. Missing the forest for the trees. *J. Sci. Educ. Technol.* **2007**, *16*, 155–169.

46. Carlone, H.B.; Haun-Frank, J.; Webb, A. Assessing equity beyond knowledge- and skills-based outcomes: A comparative ethnography of two fourth-grade reform-based science classrooms. *J. Res. Sci. Teach.* **2011**, *48*, 459–485.
47. Upadhyay, B.R. Practicing reform-based science curriculum in an urban classroom: A Hispanic elementary schools teacher's thinking and decisions. *School Sci. Math.* **2005**, *105*, 343–351.
48. Berube, C.T. Are standards preventing good teaching? *Clear. House* **2004**, *77*, 264–267.
49. Herman, J.L. Accountability and Assessment in the Service of Learning: Is Public Interest in K-12 Education Being Served? In *The Future of Test-Based Accountability*; Ryan, K.E., Shepard, L.A., Eds.; Taylor and Francis: New York, NY, USA, 2008; pp. 211–232.
50. Herman, J.L. *Moving to the Next Generation of Standards for Science: Building on Recent Practices* (CSE Tech. Rep. No. 762); University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST): Los Angeles, CA, USA, 2009.
51. Herman, J. *Making Accountability Work to Improve Student Learning* (CSE Tech. Rep. No. 649); University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST): Los Angeles, CA, USA, 2005.
52. Polikoff, M.S. Instructional alignment under No Child Left Behind. *Am. J. Educ.* **2012**, *118*, 341–368.
53. Au, W. High-stakes testing and curricular control: A qualitative metasynthesis. *Educ. Res.* **2007**, *36*, 258–267.
54. Marx, R.W.; Harris, C.J. No Child Left Behind and science education: Opportunities, challenges, and risks. *Elem. School J.* **2006**, *106*, 467–478.
55. Raudenbush, S.W. What are value-added models estimating and what does this imply for statistical practice? *J. Educ. Behav. Stat.* **2004**, *29*, 121–129.
56. Bryan, L.A.; Atwater, M.M. Teacher beliefs and cultural models: A challenge for science teacher preparation programs. *Sci. Educ.* **2002**, *86*, 821–839.
57. Goldston, D. Elementary science: Left behind? *J. Sci. Teach. Educ.* **2005**, *16*, 185–187.
58. Geier, R.; Blumenfeld, P.; Marx, R.; Krajcik, J.; Fishman, B.; Soloway, E. Standardized Test Outcomes of Urban Students Participating in Standards- and Project-based Science Curricula. In *Proceedings of the Sixth International Conference of the Learning Science*; Kafai, Y., Sandoval, W., Enyedy, N., Nixon, A., Herrera, F., Eds.; Erlbaum: Mahwah, NJ, USA, 2004; pp. 310–317.
59. Songer, N.B.; Lee, H.S.; Kam, R. Technology-rich inquiry science in urban classrooms: What are the barriers to inquiry pedagogy? *J. Res. Sci. Teach.* **2002**, *39*, 128–150.
60. Abrams, L.M.; Pedulla, J.J.; Madaus, G.F. Views from the classroom: Teacher's opinion of statewide testing programs. *Theor. Prac.* **2003**, *42*, 18–29.
61. Pedulla, J.; Abrams, L.; Madaus, G.; Russell, M.; Ramos, M.; Miao, J. *Perceived Effects of State Mandated Testing Programs on Teaching and Learning*; National Board on Educational Testing and Public Policy: Boston, MA, USA, 2003.
62. Liu, O.L.; Lee, H.-S.; Hofsetter, C.; Linn, M.C. Assessing knowledge integration in science: Construct, measures, and evidence. *Educ. Assess.* **2008**, *13*, 33–55.
63. Yeh, S.S. Test worth teaching to: Constructing state-mandated tests that emphasize critical thinking. *Educ. Res.* **2001**, *30*, 12–17.

64. Songer, N.B.; Wenk Gotwals, A. Guiding explanation construction by children at the entry points of learning progressions. *J. Res. Sci. Teach.* **2012**, *49*, 141–165.
65. Erlandson, D.A.; Harris, E.L.; Skipper, B.L.; Allen, S.D. *Doing Naturalistic Inquiry: A Guide to Methods*; Sage: London, England, 1993.
66. Guba, E.G.; Lincoln, Y.S. *Effective Evaluation*; Jossey-Bass Publishers: San Francisco, CA, USA, 1981.
67. Merriam, S.B. *Qualitative Research and Case Study Applications in Education*; Jossey-Bass: San Francisco, CA, USA, 1998.
68. Kid's Count Data Center. *Data by State*, Available online: <http://datacenter.kidscount.org/data/bystate/Default.aspx> (accessed on 10 June 2011).
69. Duncan-Andrade, J. Gangstas, wankstas, and ridas: Defining, developing, and supporting effective teachers in urban schools. *Int. J. Qual. Stud. Educ.* **2007**, *20*, 617–638.
70. Ladson-Billings, G. *The Dreamkeepers: Successful Teachers of African American Children*; Jossey-Bass Publishers: San Francisco, CA, USA, 1994.
71. Patton, M.Q. *Qualitative Research & Evaluation Methods*, 3rd ed.; Sage Publications, Inc.: Thousand Oaks, CA, USA, 2002.
72. Farr, S. *Teaching as Leadership: The Highly Effective Teacher's Guide to Closing the Achievement Gap*; Jossey-Bass: San Francisco, CA, USA, 2010.
73. Lazar, A. Literacy teachers making a difference in urban schools: A context-specific look at effective literacy teaching. *J. Read. Educ.* **2006**, *32*, 13–21.
74. Peterson, K.D.; Bennet, B.; Sherman, D.F. Themes of uncommonly successful teachers of at-risk students. *Urban Educ.* **1991**, *26*, 176–194.
75. Clark, C.M.; Peterson, P.L. Teachers' thought Processes. In *Handbook of Research on Teaching*, 3rd ed.; Wittrock, M.C., Ed.; Macmillan: New York, NY, USA, 1986; pp. 255–296.
76. Joyce, B. Toward a theory of information processing in teaching. *Educ. Res. Quart.* **1978–79**, *3*, 66–77.
77. Erickson, F. Qualitative Methods in Research on Teaching. In *Handbook of Research on Teaching*, 3rd ed.; Wittrock, M.C., Ed.; Macmillan: New York, NY, USA, 1986; pp. 119–161.
78. Creswell, J.W. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*; Sage: Los Angeles, CA, USA, 2009.
79. Freeman, M.; deMarrais, K.; Preissle, J.; Roulston, K.; St. Pierre, E.A. Standards of evidence in qualitative research: An incitement to discourse. *Educ. Res.* **2007**, *36*, 25–32.
80. NECAP. *Guide to Using the 2011 NECAP Science Reports, 2011*. Available online: <http://www.ride.ri.gov/Portals/0/Uploads/Documents/Instruction-and-Assessment-World-Class-Standards/Assessment/Assessment-Schedule-Workshops/Guide-to-Using-the-2011-NECAP-Science-Reports.pdf> (accessed on 16 October 2012).
81. NECAP. *Released Items Support Materials 2011: Grade 4 Science, 2011*. Available online: <http://www.ride.ri.gov/Portals/0/Uploads/Documents/Instruction-and-Assessment-World-Class-Standards/Assessment/NECAP/Released-Items/NECAP-2011-Gr4-Science-Released-Items.pdf> (accessed on 16 October 2012).

82. NECAP. *NECAP Assessment of Inquiry*. Available online: <http://www.ride.ri.gov/Portals/0/Uploads/Documents/Instruction-and-Assessment-World-Class-Standards/Science/NECAP-InquirySchema.pdf> (accessed on 16 October 2012).
83. NECAP. *Released Items Support Materials 2008: Grade 4 Science, 2008*. Available online: <http://www.ride.ri.gov/Portals/0/Uploads/Documents/Instruction-and-Assessment-World-Class-Standards/Assessment/NECAP/Released-Items/NECAP-2008-Gr4-ScienceReleasedItems-SupportMaterials.pdf> (accessed on 16 October 2012).
84. NECAP. *Released Items Support Materials 2009: Grade 4 Science, 2009*. Available online: <http://www.ride.ri.gov/Portals/0/Uploads/Documents/Instruction-and-Assessment-World-Class-Standards/Assessment/NECAP/Released-Items/2009-Grade-4-Science-Released-Items-Support-Materials.pdf> (accessed on 16 October 2012).
85. NECAP. *Released Items Support Materials 2010: Grade 4 Science, 2010*. Available online: <http://www.ride.ri.gov/Portals/0/Uploads/Documents/Instruction-and-Assessment-World-Class-Standards/Assessment/NECAP/Released-Items/NECAP-2010-Gr4-Science-Released-Items-V2.pdf> (accessed on 16 October 2012).
86. NECAP. *Task Booklet 2008 Grade 4 Released Science Inquiry Task: Bird Beaks and Survival, 2008*. Available online: <http://www.ride.ri.gov/Portals/0/Uploads/Documents/Instruction-and-Assessment-World-Class-Standards/Assessment/NECAP/Released-Items/NECAP-2008-Gr4-ScienceReleased-TaskBooklet.pdf> (accessed on 16 October 2012).
87. Michaels, S.; O'Connor, M.C.; Hall, M.W.; Resnick, L.B. *Accountable Talk[®] Sourcebook: For Classroom Conversation That Works*; University of Pittsburgh: Pittsburgh, PA, USA, 2010.
88. Institute for Inquiry. *Fundamentals of Inquiry*, Available online: <http://www.exploratorium.edu/ifi/workshops/fundamentals/> (accessed on 10 June 2011).
89. Full Option Science System (FOSS). *Magnetism and Electricity: Teacher Guide*; Delta Education: Nashua, NH, USA, 2005.
90. National Academy of Sciences (NAS). *Electric Circuits: Teacher's Guide*; Carolina Biological Supply Company: Burlington, NC, USA, 2004.
91. Wiggins, G.; McTighe, J. *Understanding by Design*; Association for Supervision and Curriculum Development: Alexandria, VA, USA, 2005.
92. Keys, C.W.; Kennedy, V. Understanding inquiry science teaching in context: A case study of an elementary teacher. *J. Sci. Teach. Educ.* **1999**, *10*, 315–333.
93. Kuhn, D.; Pease, M. What needs to develop in the development of inquiry skills? *Cogn. Instruct.* **2008**, *26*, 512–559.
94. Metz, K.E. On the complex relation between cognitive developmental research and children's science curricula. *Rev. Educ. Res.* **1997**, *67*, 151–163.
95. Metz, K.E. Children's understanding of scientific inquiry: Their conceptualization of uncertainty in investigations of their own design. *Cogn. Instruct.* **2004**, *22*, 219–290.
96. Johnson, D.W.; Johnson, R.T.; Stanne, M.B. *Cooperative Learning Methods: A Meta-Analysis, 2000*. Available online: <http://www.ccsstl.com/sites/default/files/Cooperative%20Learning%20Research%20.pdf> (accessed on 16 December 2012).
97. Mercer, N.; Dawes, L.; Wegerif, R.; Sams, C. Reasoning as a scientist: Ways of helping children to use language to learn science. *Br. Educ. Res. J.* **2004**, *30*, 359–377.

Appendix 1

Weekly Semi-Structured Interview Guide

1. Describe your plans for science instruction for this week.
2. Suppose I was present as you planned, what would I see you doing or how would you describe your thought process?
3. What factors influenced your decisions for this week?
4. Describe the resources that you used to assist you with your planning (*i.e.*, curricular materials, team planning, experts).
5. Describe any resources you needed that you sought out on your own.
6. Describe your feelings about your content knowledge for this week's science lessons.
7. What did or will you do to acquire the content knowledge for this week's lessons?
8. Describe the impact, if any, that the students had on your planning this week?
9. What specific aspects of science did you explicitly plan for students to experience for this week's science lessons?
10. Describe your view of the role you will take in the science classroom this week.
11. Describe your view of the role that your students will take in the science classroom this week.

Selected Emergent Questions about NECAP Testing

1. In which areas did your students make progress last year on the NECAP science test?
2. How are your plans this year different from past years as a result of the information you received from the released items or the test results?
3. Why do you take note of the NECAP science test results if they do not affect AYP?
4. How do you make time for science with the district's emphasis on reading and math?
5. You mentioned that you plan to focus this year on _____ in preparing your students for the NECAP science test? Why?
6. You mentioned that your goal is to promote your students' questioning? What is involved in planning for student questioning?
7. You mentioned that you wanted students to design their own procedures for investigations? What is involved in that planning?
8. (Lee) You explained that you focus on data collection and data representation? Why? What decisions do you make to promote your students' development of these practices?
9. What steps do you take or decisions do you make to promote students' explanations about their findings?
10. (Ann) You mentioned that your goal is for students to evaluate their data and their claims. What steps do you take to promote students' development of these practices?
11. Describe the nature of your students' conversations in science.
12. Describe the nature of your students' written conclusions.

Appendix 2

List of Open Codes

A. Planning for the teacher role

- A1 Classroom climate
- A2 Teacher questioning
- A3 Modeling
- A4 Teacher guided vs. open investigations
- A5 Teacher guided discourse
- A6 Connections to real life

B. Planning for the student role

- B1 Collaboration in teams
- B2 Student questioning
- B3 Generating testable questions
- B4 Designing investigation
- B5 Designing data recording
- B6 Representing data
- B7 Student-generated scientific discourse
- B8 Written documentation (see below)

C. Planning in response to NECAP science test

- C1 Goals for science instruction
- C2 Controlling variables
- C3 Designing a fair test
- C4 Recording data
- C5 Representing data in graph form
- C6 Interpreting graphs
- C7 Explaining reasoning
- C8 Writing conclusions
- C9 Applying findings to situations

D. Planning for science discourse

- D1 Science vocabulary
- D2 Reporting out of findings
- D3 Developing explanations based on evidence
- D4 Evaluating evidence
- D5 Listening to peer's claims
- D6 Evaluating claims
- D7 Building on other's ideas
- D8 Considering alternative explanations

E. Planning for writing

- E1 Data recording
- E2 Data representation
- E3 Summary of findings
- E4 Written claims supported by evidence
- E5 Written evaluation of evidence and/or claims

F. Planning for students with special needs

- F1 English language learners
- F2 Students with learning differences

G. Professional Development Training

- G1 Reform-based science
- G2 Literacy
- G3 Evidence-based discourse

H. Resources Available

- H1 NECAP science test released items
- H2 Mentors
- H3 Collaborating colleagues
- H4 Commercial science kits
- H5 Teacher-developed materials