Volume 13 | Number 2 | 2009

Electronic Journal of Science Education

ISSN 1087-3430

Published by Southwestern University

Electronic Journal of Science Education
Volume 13, Issue 2 (2009)
CONTENTS

Editorial:

Vygotsky and Science Education

Michael Kamen
Southwestern University

Obuchenie, a Russian word with no direct English translation, meaning both teaching and learning became one of a number of terms and concepts I have recently been grappling with. Luis Moll (1990) described *Obuchenie* as "be[ing] used for both the activities of students and teachers, implicating a double sided process of teaching/learning, a mutual transformation of teacher and student." But seeing this notion implemented in a real school in a small rural Russian city gave me access to its meaning more fully.

I had an opportunity to attend a week long "summer school" in Belaya Kalitiva, Russia (an overnight train ride south of Moscow). This institute/workshop/conference was organized by the Vygotsky Institute at the Russian State University for the Humanities and led by the Institute's director, Elena Kravtsova (Lev Vygotsky's granddaughter). It was hosted at an elementary school that is one of a number of *Golden Key* schools in Russia. These schools boast an organizational structure, curriculum, and pedagogy based on the work of Vygotsky. In addition to attending presentations from both Russian Vygotskian scholars and an international group of participants we were fortunate to observe students at the elementary school during the week-long summer school.

The experience was transformational for me. I have drawn from Vygotsky for many years in my own work for scholarship, pedagogy, and instructional methods I share with beginning teachers to support their efforts to become good science teachers. Yet this experience quickly led me to confront my own understanding of concepts such as the *Zone of Proximal Development* and introduced me to deeper meanings as well other provocative ideas.

The ongoing scholarship of Vygotsky's (and his peers and followers) theories and research, which may also be found under the rubrics of *Cultural-Historical Psychology* and *Activity Theory*, continues to hold my attention as a powerful theoretical frame to inform science education thought and practice. I am intrigued by Lois Holzman's (1997) assertion that a purely epistemological theory of learning provokes a limited view of development and in turn supports curricula and pedagogical decisions that actually limits a child's development and learning. Her argument suggests that if we limit our theoretical and research perspective to *knowledge* we are missing important aspects of development. The question becomes how we can support a child's mediation with their *cultural-historical* world through *activity* where learning is leading development.

These ideas are just touching the surface of exciting scholarly endeavors that range from research on applying Vygotskian theory to science education to new translations of Vygotsky's *unedited* works. To push the understanding of Vygotsky's work (and those who continue his scholarship) in relation to science education I have decided to dedicate the first-ever *special issue* of the Electronic Journal of Science Education to Lev Vygotsky. My friend, science educator, and Vygotskian scholar, Colette Murphy from the Queens University in Belfast, Ireland will serve as co-editor of this issue. We invite manuscripts relating or applying Vygotskian theory and research, cultural-historical (sociocultural) theory, and activity theory to science education/science teacher education issues from early childhood through to university level, including informal science and environmental education. Manuscripts reporting research, presenting theory, and/or arguing innovative perspectives will be considered for publication. See call for papers on the EJSE website for submission information. The deadline for submission is September 30, 2010.

References

Holzman, L. (1997). *Schools for Growth: Radical alternatives to current educational models*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc

Moll, Luis C. (1990). Introduction. In Luis C. Moll (Ed.), *Vygotsky and education: instructional implications and applications of Sociohistorical Psychology* (p. 24). Cambridge, UK: Cambridge University Press.

The Achievement of Student Subgroups on Science Performance Assessments in Inquiry-Based Classrooms

Jerome M. Shaw
University of California, Santa Cruz

Sam O. Nagashima
University of California Los Angeles

Abstract

This study examined student learning in science as measured by performance assessments embedded within inquiry-based units of instruction. These locally developed assessments were implemented in a consortium of districts involved in a multi-year science education reform initiative. The sample consisted of scores from 834 fifth grade students on three performance assessments given in a participating district's 14 elementary schools during the 2004-2005 school year. District-provided data permitted disaggregation of student scores by ethnicity, gender, and socioeconomic status as well as identification as English Learner, Gifted and Talented, and Special Education. Using mean scores as the basis for comparison, results showed the majority of students achieving at the proficient level as defined by initiative-developed rubrics. Statistical analyses indicated significant underperformance on one or more of the assessments by Blacks, Hispanics, low socioeconomic status students, males, non-Gifted, and Special Education students. Depending on the performance assessment and the student subgroup, potential factors related to performance include science discipline and access to economics-related resources (e.g., computers). This study is noteworthy for its comprehensiveness and the nuanced understandings it brings to previously documented achievement gaps.

Correspondence should be addressed to Jerome M. Shaw at jmlshaw@ucsc.edu, or Sam O. Nagashima at samn@ucla.edu

Introduction

Measuring student achievement in science recently has garnered greater attention across the United States. This increase is due in part to the onset of the No Child Left Behind Act's (NCLB) requirement of statewide testing in science starting with the 2007-2008 school year (NCLB, 2002). Federal law now requires local education agencies to annually assess their students' learning in core academic subjects such as science in relation to content that is specified in state-sanctioned standards (U.S. Department of Education, 2002). Said standards routinely include concepts from multiple scientific disciplines such as biology and physics as well as process skills such as designing and conducting experiments.

While NCLB-mandated content standards provide broad guidance as to what should be assessed, state policymakers are left to decide how. Among the many issues to consider when choosing an assessment methodology are feasibility and compatibility. With respect to the former, factors such as time, material requirements and associated costs make selected response testing (e.g., multiple choice) more attractive than approaches such as performance assessment – simply defined as "assessments that allow students to demonstrate their understandings and skills… as they perform a certain activity" (National Research Council [NRC], 2001, p. 31) – which can be difficult as well as costly to develop and implement (Baker, 1997). Science performance assessments in particular may be up to 100 times more expensive than multiple-choice tests (Stecher, 1995) and three times more expensive than open-ended writing assessments (Stecher & Klein, 1997).

Perhaps more important than monetary cost is consideration of an assessment practice's degree of compatibility with the standards it is designed to measure. As the *National Science Education Standards* (NSES) state, "assessments provide an operational definition of standards, in that they define in measurable terms what teachers should teach and students should learn" (NRC, 1996, pp. 5-6). In clarifying this notion of teaching and learning, the NSES and its companion volume on classroom assessment (NRC, 2001), stress that a distinguishing feature of reform science teaching is its focus on engaging students in "active and extended scientific inquiry" (NRC 1996, p. 52). Such inquiry-based instruction leads to the concomitant building of content knowledge and development of process skills.

In theory, more so than selected response methods, performance assessments are especially well suited to measure the complex mix of conceptual understandings and science process skills associated with inquiry-based instruction (NRC, 2001, 2005; Shavelson, Baxter, & Pine, 1991). In practice, most NCLB-compliant tests rely heavily on selected response formats. However, the authors of *Inquiry and the National Science Education Standards* (NRC, 2000) fault multiple-choice tests for being too broad in coverage and focused on recognition and recall of facts, attributes which predispose them to pose "a serious obstacle" to inquiry-based teaching (p. 75). Thus, the appropriateness of selected response assessment for measuring the achievement of students taught inquiry-based science is questionable.

In contrast to the varying approaches to assessing student learning, proponents of current reforms are consistently resolute in asserting that inquiry-based science education is for all students (American Association for the Advancement of Science, 1989; NRC, 1996, 2000). Reform efforts funded by agencies such as the National Science Foundation work to realize this vision in part by aligning multiple elements of the educational enterprise including curriculum, instruction and assessment. One aspect of this alignment is the linking of inquiry-based curriculum and instruction with assessment that is performance-based.

Congruence between inquiry-based science instruction and performance assessment notwithstanding, several studies document student achievement in inquiry-based science classrooms solely using traditional forms of assessment, in part due to the feasibility issues discussed above (Amaral, Garrison, & Klentschy, 2002; Geier et al., 2008; Johnson, Kahle, & Fargo, 2007; Lee, Deaktor, Hart, Cuevas, & Enders, 2005; Lynch, Kuipers, Pyke, & Szesze, 2005). Other studies of student learning in hands-on science classrooms do utilize performance assessments, either exclusively (Cuevas, Lee, Hart, & Deaktor, 2005; Lee, Buxton, Lewis, & LeRoy, 2006; Shaw, 1997) or in conjunction with traditional assessments (Pine et al., 2006). Still other studies use performance assessment to evaluate student learning in science without specifying the instructional context (Klein et al., 1997; Lawrenz, Huffman, & Welch, 2001).

This sparse literature base provides some indication of the extent to which the inclusive goals of science education reform are realized in terms of the performance of diverse student subgroups. While mixed, the combined findings from these various studies indicate achievement gaps by gender (females outscoring males), ethnicity (Whites outscoring non-whites), socioeconomic status or SES (high SES outscoring low SES), "giftedness" (Gifted students outscoring Non-Gifted), English proficiency (Non-English Learners outscoring English Learners). The studies noting students' native language show no clear pattern. These findings, accompanied with source information, are presented in Table I. Worth noting is that, of the various student subgroups reported on, there is a distinct omission of findings for students designated as Special Education, whether with traditional or performance assessments. Also absent from the literature are findings on Gifted students based on performance assessment. Finally, the most student demographic subgroups reported in any of these recent studies is five by Lee and colleagues in 2005.

Table I

Synthesized findings of student subgroup performance in inquiry-based science classrooms

	ASSESSMENT TYPE	
SUBGROUP	TRADITIONAL	PERFORMANCE
Gender	• Girls > Boys (3*, *6**) • Girls = Boys (4, 7, 10) • Boys > Girls (*5**)	• Girls > Boys (*5**, *6*, 8, 10*) • Girls = Boys (7, 10)
English Learner	• Non-EL > EL (1*, 7*)	• Exited EL > Non-EL (8)
Ethnicity	• Whites > Non-White (*5**, *6**, 7*, 9) • Whites = Non-Whites (4[1])	• Whites > Non-White (*5**, *6**, 7*, 9)
Gifted	• Gifted > Non-Gifted (7*)	• Not addressed
Native Language	• English > Non-English (7)	• English = Spanish (2) • Spanish > English (8)
Socioeconomic Status	• High SES > Low SES (7*, 9, 10*)	• High SES > Low SES (9, 10*)

Notes:

* = Finding was statistically significant

Italics = unspecified instructional context (all others are inquiry-based)

Code for Studies Cited:

1. Amaral et al. (2002)
2. Cuevas et al. (2005)
3. Geier et al. (2005)
4. Johnson et al. (2007)
5. *Klein et al. (1997)*
6. *Lawrenz et al. (2001)*
7. Lee et al. (2005)
8. Lee et al. (2006)
9. Lynch et al. (2005)
10. Pine et al. (2006)

Research Purpose and Questions

This study was undertaken to broaden the empirical literature base on student achievement in science classrooms as measured by performance assessment. We present an uncommonly comprehensive set of findings by addressing six categories of student demographic subgroups, including Gifted and Special Education, in a single study. Specifically, with respect to fifth grade student scores on three performance assessments

[1] Whites significantly outperformed Non-Whites in non-inquiry-based classrooms

implemented in inquiry-based science classrooms, this study investigated the following questions:

1. What are the patterns of performance for all students and specific subgroups of students? (e.g., comparison of mean scores for students of different ethnicities)
2. Are there statistically significant differences in the performance of student demographic subgroups?

Research Design

Context

The data for this study were drawn from one of several districts that participated in a recently completed multi-year, NSF-funded science education reform initiative known as STEP-uP (Science Teacher Enhancement Program unifying the Pikes Peak region – www.stepupscience.org). STEP-uP's efforts to improve student learning included the development of performance assessments that were embedded within inquiry-based curriculum units taught at grades kindergarten through five. As part of their involvement in STEP-uP, teachers in participating districts engaged in professional development on the science curriculum units and their associated assessments.

The focal district for this study, herein referred to as the Abacus School District, was chosen for its high degree of diverse students. We chose to focus on the fifth grade level in order to reduce potential interference from student lack of familiarity with performance assessment; in other words, to avoid an issue with "opportunity to test" (Shaw, 1997). Assuming prior instruction in the participating districts, fifth grade students were more likely to have previously encountered a STEP-uP science performance assessment than those in lower grades. We chose data from the 2004-2005 school year as it was the first time that all three performance assessments for the fifth grade science units were implemented.

Curriculum

STEP-uP affiliated districts utilize a carefully selected mosaic of research-based science curriculum units from a variety of sources including FOSS, Insights, and Science and Technology for Children. At the time of this study, schools in the Abacus School District taught three science units at the 5th grade level: <u>Ecosystems</u>, <u>Food Chemistry</u>, and <u>Microworlds</u>. All of these units are from the Science and Technology for Children (STC) curriculum developed by the National Science Resources Center (NSRC) at the Smithsonian Institute, and supported by the National Academy of Sciences (NSRC, 1991, 1994, 1996). STC is an inquiry-based curriculum designed to meet the NSES and shown to produce meaningful learning gains with culturally and linguistically diverse students (Amaral et al., 2002).

Assessment

The measures of student learning used in this study were the three STEP-uP-developed performance assessments that correspond to the three science units taught at the fifth grade level: Ecosystems, Food Chemistry, and Microworlds (we use these titles to refer to the assessments as well). These assessments, which are based on and serve as replacements for lessons already present in the units, engage students in applying understandings and skills from prior lessons to new situations (e.g., for Food Chemistry, determining the nutritional value of previously untested snack foods using techniques and knowledge gained beforehand). While the assessments for Food Chemistry (which focuses on the selection of a nutritious snack) and Microworlds (which involves the observation of live microbes) replace end-of-unit lessons and serve as culminating activities, the assessment for Ecosystems (which includes the study of food webs) begins midway and continues to the end of the unit.

The assessments were developed by Design Teams composed of two to three classroom teachers with prior experience teaching the particular unit and a university scientist knowledgeable in the unit's science content. Design Team members were enrolled in a college level course led by an assessment development expert. They created the assessments and their accompanying manuals as part of the course requirements. Following initial development, the assessments underwent an iterative review and revision process that included pilot and field-testing in project-affiliated schools spread across all five STEP-uP districts. Efforts were made to have test sites reflect the student diversity of the participating districts in terms of ethnicity, socioeconomic status, Special Education, and English Learners. The full development cycle for each assessment spanned a three-year process of initial design (year one), pilot/field-testing (year two), and implementation (year three). Further details regarding the STEP-uP assessment development process can be found in Kuerbis and Mooney (2008).

For each assessment, the development process culminated with the creation of a manual containing information such as administration guidelines and correlations to national and state science content standards. Reading across manuals reveals a common focus on the science process skills of understanding scientific investigation and design as well as appropriate communication of the results from scientific investigations (STEP-uP, 2003; STEP-uP, 2004; STEP-uP, 2005).

The administration guidelines specify logistics such as requisite materials, student grouping (e.g., pair or small group), and the number, focus and suggested time length of class sessions. Expected completion times for this group of assessments range from two to seven hours over two to four sessions, each of which may range from 30 minutes to an hour or more. There is also variation with respect to the final product and/or performance on which students will be judged. For example, with Ecosystems students give an oral presentation explaining their poster on a particular ecosystem while Microworlds calls for submission of a written report.

The assessment manuals also contain black line masters of handouts needed for students to engage in the assessment. These include task sheets with instructions and rubrics for scoring student performance as well as previously scored samples of student work. Each assessment is accompanied by at least two rubrics. Returning to the example of Ecosystems, there are separate rubrics for the oral presentation and the poster on which the presentation is based. These and other key features of the assessments are provided in Table II.

STEP-up performance tasks incorporate self-assessment. Teachers are instructed to have students evaluate their own performance using slightly modified versions of the same rubrics used by teachers themselves. Teachers introduce these rubrics to the students at various stages in the assessment process (e.g., prior to creating the Ecosystems poster and while students prepare to give their oral presentation) so that the rubrics may guide student work. As part of completing an assessment, students are expected to rate their own performance using the rubrics' criteria.

Table II
Key features of the three 5th grade science performance assessments

	ECOSYSTEMS	FOOD CHEMISTRY	MICROWORLDS
Task	Research (e.g., using library books and the Internet) an ecosystem and create a poster that presents the relationships within it	Conduct physical and chemical tests on a variety of snack foods to determine the lack or presence of specific nutrients	Examine various water samples with a microscope and determine which is safest to drink
Primary Content Focus	Life science – relationships in ecosystems	Physical science – nutrient composition of foods	Life science – structure and function of microorganisms
Process Focus	Document-based research	Laboratory tests and data collection	Microscope use and observations
Rubrics (Assessment Foci)	• Ecosystem Poster • Oral Presentation	• Testing Chart • Nutrient Testing • Oral Interview	• Lab Work • Lab Report
Student Grouping	Pairs	Small Groups	Pairs
Class Sessions (Total time)	3 (5-7 hours)	4 (3-3.5 hours)	2 (2-3 hours)

Students

The 834 unique fifth grade students in the study sample were from 39 classrooms in the 14 elementary schools of the Abacus School District. District-provided data included information on individual students' membership in multiple demographic categories such as those required for disaggregation by NCLB. Listed in alphabetical order, the six student categories with their associated subgroups used in this study are as follows: English Learner, ethnicity (American Indian/Alaskan Native, Asian, Black, Hispanic, and White), gender (male, female), Gifted and Talented, socio-economic status or SES (based on status as a recipient of free or reduced lunch), and Special Education. While the categories English Learner and Special Education had multiple subgroups (e.g., limited English proficient and non-English proficient for the former; autism and physical disability for the latter), the small number of individuals in these categories warranted their being collapsed into single variables for this study. Deeper analysis of the English Learner data is reported elsewhere (Shaw, 2009).

Scores

Student scores were derived from the application of STEP-uP developed rubrics to student products or performances such as posters and oral presentations. Teachers scored their own students' responses and submitted them to the school district office. The administration guidelines instruct teachers not to submit scores for students who missed 50% or more of the instruction associated with a particular science unit. Without specifying a particular approach (such as averaging), the guidelines likewise direct teachers to give students one overall score based on their respective scores for an assessment's two or three rubrics. Thus, individual student scores were in the form of single digit numbers on a four-point scale on which 4 = Advanced, 3 = Proficient, 2 = Partially Proficient, and 1 = Unsatisfactory. Although the rubrics contain task-specific criteria, these four performance levels are common to all three assessments.

The Abacus School District provided a combined total of 2,155 individual scores (one score per assessment per student, maximum three scores per student) that represent the full complement of teacher-submitted data from the 2004-2005 school year. Not all students completed all the assessments. For the study sample (n=834), 107 (12.8%) completed only one assessment, 136 (16.3%) completed two assessments, and the remaining 591 (70.9%) completed all three assessments. With respect to the individual assessments, completion rates are 727 (87.2%) for Ecosystems, 694 (83.2%) for Food Chemistry, and 731 (87.6%) for Microworlds. Completion rates for the subgroups in the study are shown in Table III.

Method

This exploratory, post-hoc investigation[2] employed two levels of analysis: descriptive comparisons of overall and student subgroup performance, and significance testing of subgroup differences. Raw score means on each assessment were the basis for the former while the latter were conducted using z-scores. Given their connection to rubric-defined performance levels, raw scores provide a "readily understood reference point" from which to understand the comparisons (Hoover, 1984, p. 13). Conversion of raw scores to z-scores for significance testing is an accepted practice for studies of this nature (see for example, Klein et al., 1997, and Pine et al., 2006) with sound psychometric backing (Binder, 1984; Jaccard & Wan, 1996; Kim, 1975; Labovitz, 1967, 1970; Zumbo & Zimmerman, 1993).

Multiple regression analyses were conducted to determine the existence of significant differences between the performances of the student subgroups in our sample. A unique linear regression analysis was run for each of the three assessments with white females who are non-English Learners, non-Free/Reduced Lunch, non-Special Education,

[2] This study was conducted near the end of funding for the focal project and was not part of that project's original design.

and non-Gifted and Talented serving as the basic comparison group. For each of these analyses, student demographics were the independent variables (e.g., Ethnicity, Gender, SES) with assessment scores (i.e., separate values for Ecosystems, Food Chemistry, and Microworlds) as the dependent variable. Given their relatively large sample sizes, additional analyses were run for the Ethnicity subgroups (e.g., Black, Hispanic). Dummy coding was used where each subgroup was given its own variable with the exception of White, which served as the comparison group. All other student demographic variables (i.e., Gifted and Talented, Male, English Learner, SES, and Special Education,) were dichotomously coded, 0 = non-member, 1 = member.

Table III
Completion rates for student groups on the three assessments

	Ecosystems	Food Chemistry	Microworlds
ENGLISH LEARNER			
English Learners	51	61	62
ETHNICITY			
American Indian/Alaskan Native	16	17	17
Asian	28	30	29
Black	199	178	196
Hispanic	247	242	250
White	235	225	237
GENDER			
Female	365	351	368
Male	362	343	363
GIFTED AND TALENTED			
Gifted	34	30	41
SOCIO-ECONOMIC STATUS			
Free/Reduced Lunch	488	467	502
SPECIAL EDUCATION			
Special Educational Needs	74	68	75
Total Sample	727	694	731

Note. The total sample includes 834 unique students.

Findings

We present our findings using the two research questions as frames of reference. First, we provide comparisons of mean scores for all students as well as student demographic subgroups on each of the three assessments (Question 1). These findings appear in order from macro to micro in terms of the level of student groupings: total sample, reference groups (e.g., English Learner and Non-English Learner), and ethnicity subgroups. These descriptive comparisons are followed by a presentation of findings from the tests of significance for differences in subgroup performance (Question 2). This final section, titled multiple regression analyses, includes effect sizes and the degree to which student level variables explain the variance in the scores of each of the three assessments.

Total Sample Comparisons

With a variance of only .01 of a point on a scale of 1-4, mean scores on all three assessments were essentially identical for students as a whole. Mean scores for the total sample on the three assessments were: 2.80 ($SD = .837$) for Ecosystems, 2.81 ($SD = .858$) for Food Chemistry, and 2.80 ($SD = .804$) for Microworlds. We refer to this result as the "homogeneity of means" pattern, whose "universal mean" is taken to be 2.8 (see Table IV).

In terms of the levels used in the project-developed rubrics, overall student performance on the three assessments was in the "Proficient" range (2.50 – 3.49). There were no mean scores at the "Advanced" (3.50 – 4.0) or "Unsatisfactory" levels (0 – 1.49), and two mean scores at the "Partially Proficient" level (1.50 – 2.49): American Indian/Alaskan Native and Special Education on Food Chemistry, 2.47 and 2.22, (see Table V and Table IV, respectively).

Table IV
Mean scores and standard deviations for student groups on the three assessments

	Ecosystems M (S.D.)	Food Chemistry M (S.D.)	Microworlds M (S.D.)
Total Sample	2.80 (.837)	2.81 (.858)	2.80 (.804)
ENGLISH LEARNER			
Non-English Learner	2.80 (.844)	2.82 (.861)	2.81 (.806)
English Learner	2.82 (.740)	2.80 (.833)	2.69 (.781)
Difference	*0.02*	*0.02*	*0.12*
ETHNICITY			
White	2.87 (.754)	2.95 (.754)	2.89 (.766)
Non-White	2.77 (.873)	2.76 (.898)	2.75 (.818)
Difference	*0.10*	*0.19*	*0.14*
GENDER			
Female	2.92 (.808)	2.99 (.832)	2.92 (.810)
Male	2.69 (.851)	2.64 (.849)	2.67 (.779)
Difference	*0.23*	*0.35*	*0.25*
GIFTED AND TALENTED			
Non-Gifted	2.78 (.840)	2.79 (.856)	2.76 (.789)
Gifted	3.29 (.579)	3.47 (.629)	3.46 (.602)
Difference	*0.51*	*0.68*	*0.70*
SOCIO-ECONOMIC STATUS			
Non-Free/Reduced Lunch	2.92 (.747)	2.91 (.826)	2.92 (.810)
Free/Reduced Lunch	2.75 (.872)	2.78 (.870)	2.75 (.796)
Difference	*0.17*	*0.13*	*0.17*
SPECIAL EDUCATION			
Non-Special Educational Needs	2.85 (.813)	2.89 (.815)	2.87 (.762)
Special Educational Needs	2.39 (.934)	2.22 (1.01)	2.21 (.920)
Difference	*0.46*	*0.67*	*0.66*

Note. Maximum score = 4.

Other patterns worth noting are that the mean scores on all three assessments for females (2.92, 2.99 and 2.92, respectively for Ecosystems, Food Chemistry and Microworlds) and for Gifted and Talented students (3.29, 3.47 and 3.46, respectively for Ecosystems, Food Chemistry and Microworlds) were consistently above those for the total sample. The reverse was true for males (2.69, 2.64 and 2.67, respectively for Ecosystems, Food Chemistry and Microworlds) and students classified as Special Education (2.39, 2.22 and 2.21, respectively for Ecosystems, Food Chemistry and Microworlds). Mean scores for Non-white students as a group (2.77, 2.76 and 2.75, respectively for Ecosystems, Food Chemistry and Microworlds) reflect the homogeneity of means pattern.

Reference Group Comparison

The subgroups and corresponding reference groups presented here are as follows (SUBGROUP / Reference Group): ENGLISH LEARNER / Non-English Learner, ETHNICITY / White, GENDER / Female, GIFTED AND TALENTED / Non-Gifted, SOCIOECONOMIC-STATUS / Non-Free/Reduced Lunch, SPECIAL EDUCATION / Non-Special Education (see Table IV). With the exception of the Gifted and Talented subgroup, each reference group outperformed its counterpart in nearly all cases. For example, mean scores on all three assessments for females were consistently above those for males (2.92:2.69, 2.99:2.64 and 2.92:2.67, respectively for Ecosystems, Food Chemistry and Microworlds). Although slight, the lone departure from this pattern was the mean score for English Learners which was marginally higher than that for Non-English Learners on the Ecosystems assessment only (2.80:2.82, respectively).

Ethnicity Subgroup Comparisons

Listed in alphabetical order, subgroups within the Ethnicity category are American Indian/Alaskan Native, Asian, Black, Hispanic, and White (see Table V). Ethnicity subgroup mean scores range from a high of 3.23 (Asians on Food Chemistry) to a low of 2.47 (American Indian/Alaskan Native on Food Chemistry). From high to low, the pattern on Ecosystems and Microworlds was: Asian, White, Hispanic, Black, American Indian/Alaskan Native. Blacks and Hispanics reversed their rankings on Food Chemistry, resulting in the following pattern: Asian, White, Black, Hispanic, American Indian/Alaskan Native.

Table V
Mean scores and standard deviations for Ethnicity subgroups on the three assessments

	Ecosystems M (S.D.)	Food Chemistry M (S.D.)	Microworlds M (S.D.)
ETHNICITY			
American Indian/Alaskan Native	2.50 (1.030)	2.47 (0.874)	2.53 (0.800)
Asian	3.14 (0.705)	3.23 (0.568)	3.17 (0.602)
Black	2.68 (0.972)	2.74 (0.864)	2.65 (0.936)
Hispanic	2.82 (0.778)	2.73 (0.938)	2.80 (0.719)
White	2.87 (0.754)	2.95 (0.754)	2.89 (0.766)
Total Sample	2.80 (0.837)	2.81 (0.858)	2.80 (0.804)

The reference group for all Ethnicity subgroups is White. Mean scores for Whites closely resemble the homogeneity of means pattern: 2.87/2.80, 2.95/2.82, 2.89/2.80, respectively, for Ecosystems, Food Chemistry and Microworlds. Whites were outperformed by Asians on all three assessments (2.87:3.14 / Ecosystems, 2.95:3.23 / Food Chemistry, 2.89:3.12 / Microworlds). Conversely, Whites outperformed all other Ethnicity subgroups on all three assessments.

Multiple Regression Analyses

Individual multiple regression analyses were conducted using z-scores as outcome variables and student background demographic information as the predictor variable to identify statistically significant differences in group performance. Results of theses analyses are presented in Tables VIa-c. On each of the three assessments, four to five subgroups showed statistically significant results. Common to all three assessments was the pattern of underperformance by males and Special Education students and overperformance by Gifted students. Assessment-unique instances of underperformance were low SES students on Ecosystems, Blacks and Hispanics on Food Chemistry, and Blacks on Microworlds.

Table VIa
Summary of regression coefficients for Ecosystems assessment

	ECOSYSTEMS		
Variables	B	SE B	β
Constant	.296	.085	
English Learner	-.006	.147	-.002
American Ind./Alaskan Ntv.	-.323	.252	-.047
Asian	.285	.194	.055
Black	-.146	.094	-.065
Hispanic	-.003	.095	-.001
Male	-.224	.072	-.112**
Gifted	.512	.177	.104**
SES	-.187	.078	-.088*
Special Education	-.468	.120	-.141***

*p<.05, **p < .01, ***p< .001

Table VIb
Summary of regression coefficients for Food Chemistry assessment

	FOOD CHEMISTRY		
Variables	B	SE B	β
Constant	.414	.086	
English Learner	.071	.137	.020
American Ind./Alaskan Ntv.	-.372	.239	-.058
Asian	.256	.184	.052
Black	-.199	.095	-.087*
Hispanic	-.274	.094	-.131**
Male	-.346	.072	-.173***
Gifted	.641	.177	.130***
SES	-.097	.078	-.045
Special Education	-.688	.122	-.205***

*p<.05, **p < .01, ***p< .001

With respect to gender, females outperformed their male counterparts by an average of .271 of a point on the three assessments: .224 on Ecosystems, .346 on Food Chemistry, and .223 on Microworlds. These differences were statistically significant at the p<.01 level for Ecosystems and at the p<.001 level for Food Chemistry and Microworlds.

Regarding ethnicity, two groups underperformed relative to their white counterparts: Blacks underperformed by -.199 on Food Chemistry and -.241 on Microworlds, and Hispanics by -.274 on Food Chemistry. These differences were significant at the p<.05, p<.001, and p<.01 levels, respectively. The Ecosystems assessment showed no significant difference in student performance based solely on ethnicity.

Table VIc
Summary of regression coefficients for Microworlds assessment

	MICROWORLDS		
Variables	B	SE B	β
Constant	.315	.083	
English Learner	-.156	.134	-.043
American Ind./Alaskan Ntv.	-.348	.237	-.052
Asian	.245	.180	.049
Black	-.241	.091	-.107***
Hispanic	-.072	.091	-.035
Male	-.243	.069	-.122***
Gifted	.718	.153	.164***
SES	-.141	.077	-.065
Special Education	-.701	.114	-.215***

*p<.05, **p < .01, ***p< .001

Low SES students underperformed relative to their high SES peers by -.187 on Ecosystems. This difference was significant at the p<.05 level. They also underperformed on Food Chemistry and Microworlds, however these differences were not significant.

Students classified as Special Education underperformed relative to their non-Special Education counterparts with a difference of -.468 on Ecosystems, -.688 on Food Chemistry, and -.701 on Microworlds, for an average of -.619. Each of these differences was significant at the p<.001 level.

Conversely, students classified as Gifted and Talented outperformed non-Gifted and Talented students an average of .624 on all three assessments: .512 on Ecosystems, .641 on Food Chemistry, and .718, on Microworlds. These differences were significant at the p<.01, p<.001, and p<.001 levels, respectively.

Overall, student level variables explained only a small proportion of variance in the scores for all three assessments: Ecosystems, $R^2 = .062$; Food Chemistry, $R^2 = .120$; and Microworlds, $R^2 = .127$. In general, less than 12% of the total variability in student scores is accounted for by student level variables. Estimates of effect size (f^2) suggest marginal effect due to English Learner status alone and a very small effect due to Ethnicity alone (see Table VII).

Table VII
Effect sizes of models

	ECOSYSTEMS	
	Adj. R^2	Effect size (f^2)
Complete Model[a]	.062	.066
Ethnicity Only[b]	.012	.012
English Learner Only[c]	.000[d]	-
	FOOD CHEMISTRY	
	Adj. R^2	Effect size (f^2)
Complete Model[a]	.120	.136
Ethnicity Only[b]	.026	.027
English Learner Only[c]	.000d	-
	MICROWORLDS	
	Adj. R^2	Effect size (f^2)
Complete Model[a]	.127	.138
Ethnicity Only[b]	.021	.022
English Learner Only[c]	.016	.016

Note. Effect sizes (f^2) of 0.02, 0.15, and 0.35 are considered small, medium, and large, respectively (Cohen, 1988).

[a]Includes all student level demographic variables (English Learner, Ethnicity, Gender, Gifted and Talented, SES, and Special Education)

[b]Includes only Ethnicity variables (American Indian/Alaskan Native, Asian, Black, Hispanic, and White)

[c]Includes only English Learner variables (EXIT, LEP, NEP)

[d]Models non-significant at $p < .01$

Discussion and Implications

The purpose of this exploratory study was to shed light on student learning in inquiry-based science classrooms. Given this context, our post-hoc analysis used results from three locally developed, curriculum-embedded, fifth grade performance assessments to inform an understanding of the achievement of multiple NCLB-accountability related student subgroups. Within these parameters, our research questions essentially ask: To what degree are students learning science?, and, Are there appreciable differences in the level of science learning based on group affiliation? As seen in the previous section and discussed below, the answers are mixed. Interpretations of student scores analyzed in this study need to be made with at least two considerations in mind. Although reported for individual students, the scores are (a) based on collaborative work (i.e., students worked in pairs or small groups on the assessments), and (b) represent teachers' appraisal of student overall proficiency on the various constructs measured (through unspecified processes, teachers' global scores were based on the two or three rubric-based scores students received per assessment). Therefore, the scores are relevant for discussing groups of students, not individuals.

This study is limited in scope; the data come from only one of the initiative's participating districts for one school year at one grade level. Future studies can address this shortcoming by expanding the data set along each of those dimensions (i.e., multiple districts, school years, and grade levels). An additional track along which to amplify the current study's scope would be to compare findings from the performance assessments with those from other measures of science achievement. A logical candidate for such a comparison is student scores on statewide science tests. While such a test was not in operation at the time of the study, one has been implemented since. Studies including scores from the two sources would need to consider factors such as the comparability or lack thereof in the constructs measured by each assessment.

An additional concern is the lack of comprehensive validity evidence to support the interpretation of scores from the assessments. In his discussion of validity issues pertinent to performance assessments, Messick (1996) notes the importance of "transparency," meaning that "the performance standards are understood and facilitate learning" (p. 13). This issue is partially addressed by the assessments' administration guidelines that call for teachers to share and discuss the scoring guidelines or rubrics with the students before, during and after the assessment. However, the degree to which students understand the rubrics and such teacher-student interactions facilitate learning is beyond the purview of this study. Moreover, evidence stands to be gathered along the six aspects of Messick's "unified" concept of construct validity: content, substantive, structural, generalizability, external, consequential (1996, p. 7). While an ongoing and evolving process, obtaining such evidence is important if the assessments are to be used past the existence of STEP-uP and outside the project's geographic boundaries.

Bearing those factors in mind and turning to our results, we found that, in broad terms, the universal mean of 2.8 – or "Proficient" when rounded to 3.0 – indicates that students as a whole exhibited desirable levels of science knowledge and skills. It is encouraging to note that this level of achievement held constant on all three assessments, indicating comparable performance across the different content, process, and assessment foci (see Table II). For example, students as a whole were proficient at demonstrating knowledge of ecosystems, nutrition, and microorganisms through graphic, oral, and written means, respectively.

Parsing the total sample, we uncovered patterns of underperformance for specific student subgroups on each assessment (see Table VIII). We applied two criteria to put these findings into perspective. First, we identified those differences that were statistically significant. Second, we invoked a criterion of "practical" significance meaning that, following standard conventions for rounding to the nearest whole number, the mean scores of comparison groups translated to different levels of performance on the rubric. It should be noted that there are no student groups that meet the practical significance criterion that do not also meet the criterion for statistical significance. Thus, applying them in the stated order does not eliminate any group from being identified as underperforming.

Application of the above two criteria yields a more focused appraisal of student subgroup underperformance. As shown in Table VIII, statistically significant differences were observed with respect to gender, Gifted, and Special Education across all three assessments. In addition, assessment-specific underperformance (discussed below) was as follows: low SES students on Ecosystems, Black and Hispanic students on Food Chemistry, and Black students on Microworlds. However, rounding student group mean scores to the nearest whole number and applying the practical significance criterion narrowed the field of underperformers to non-Gifted and Special Education students. These were the only two groups meeting both significance criteria and they did so on all three assessments. In relative terms, the degree of practical significance was not great – only one rubric level (i.e., Proficient versus Partially Proficient, the next lowest level). Separation by non-adjacent levels (e.g., Proficient vs. Unsatisfactory) would present cause for greater concern.

Table VIII
Significant underperformance by student groups on the three assessments

ECOSYSTEMS	FOOD CHEMISTRY	MICROWORLDS
---	Black	Black
---	Hispanic	---
Male	Male	Male
Non-Gifted	*Non-Gifted*	*Non-Gifted*
Low SES	---	---
Special Education	*Special Education*	*Special Education*

Note. All student groups listed showed statistically significant differences from regression analyses (see Table VI). Groups in italics also showed practical significance (defined as having a rubric-based performance level lower than the corresponding reference group).

Our documentation of the underperformance of non-Gifted students and Special Education students as measured by performance assessments in science classrooms, while new to the literature, is not surprising. It is likely attributable to the achievement disparities inherent in the definition of those categories. Nevertheless, studies could be undertaken to discern whether or not performance assessments can enlighten our understanding of the nature of these gaps and how they might be addressed in inquiry-based science classrooms.

Significance ratings aside, important nuances are present in the achievement patterns observed in this study. Many of our findings corroborate those of prior studies. The over-performance of females, Whites, and high SES students in relation to their respective counterparts are reflective of findings by Geier and colleagues (2005), Johnson and colleagues (2007), and Lee and colleagues (2006). Conversely, the lack of difference in performance between English Learners and Non-English Learners runs counter to results reported by Amaral and colleagues (2002) as well as Lee and colleagues (2005). The small size of the English Learner sample in our study limits the power of this finding. Future studies with larger populations are needed to determine the veracity of this claim. While not available in our data set, analyses incorporating English Learner's native language (see Cuevas et al., 2005 and Lee et al., 2006) would further clarify the nature of the complex relationship between language proficiency and student achievement.

Likewise contrary to published results (Lee et al., 2005, Lynch et al., 2005, and Pine et al., 2006) is the lack of significant performance differences between high and low SES students on two of the three performance assessments in our study. With counts in the hundreds (low SES close to 500 and high SES near 200), small sample size does not appear to be an issue here. The fact that a statistically significant difference arose only on the Ecosystems assessment may be related to the nature of that task. Given the long-term, document-based research focus of the task (conceivably including out of school time), student performance arguably might be influenced by economic factors such as the quantity and quality of relevant library reference materials as well as access to computers and the Internet. In contrast, the other two assessments are dependent on resources previously used during instruction (e.g., nutrient testing equipment for Food Chemistry and microscopes for Microworlds). Further study noting the particular location of different SES students (e.g., high SES versus low SES schools) may help clarify this ambiguous finding.

Our findings in relation to ethnicity raise similar assessment-specific concerns. Except for the performance of American Indians/Alaskan Natives on Food Chemistry (mean -2.47 equivalent to "Partially Proficient"), all Ethnicity subgroups performed at the "Proficient" level (i.e., means within the range of 2.5 – 3.4). In comparison to Whites, statistically but not practically significant differences were found for Blacks and Hispanics on Food Chemistry and for Blacks alone on Microworlds (see Table VIII). It is worth investigating whether or not the design of or scoring systems for the Food Chemistry and Microworlds assessments somehow disadvantage these subgroups. Given the wide variability of individuals within ethnic groups it is unproductive to speculate on potential explanations for these observed differences. Research employing techniques such as focus groups or think aloud protocols have the potential to provide insights on these important yet perplexing results (see for example, Martiniello, 2008).

More sense can be made of the findings in relation to gender. Gender gaps have been shown to be sensitive to assessment type and content orientation. While the general pattern is one of girls outperforming boys, Klein and colleagues (1997) found that boys outperformed girls on a multiple-choice test. However, that same study found that girls outperformed boys on performance assessments, a pattern our findings uphold. In a study using scores from four different performance assessments, Pine and colleagues (2006) found comparable gender performance on physical science tasks while girls outperformed boys on the one life science task. However, in our study girls outperformed boys on both life (Ecosystems and Microworlds) and physical science (Food Chemistry) tasks. This apparent contradiction might be explained by the strong connection of the particular physical science content to the life sciences on the Food Chemistry assessment (the chemical determination of nutrients in food was tied directly to nutritional value for human consumption). Future research should explore the persistence of this discipline-specific bias with performance assessments. Bearing in mind that different assessment formats measure different competencies (Lawrenz et al., 2001; Shavelson et al., 1991), careful attention to the particular constructs measured by the assessments yielding these results would be a necessary step to make further sense of these outcomes.

In sum, our findings bring greater comprehensiveness and additional insights to the understanding of student performance in inquiry-based science classrooms. This study is noteworthy for its inclusion of six student demographic subgroup variables (more than any recent study on the topic) and the nuanced understandings it brings to previously documented achievement gaps, gender in particular. While not intending to provide definitive answers or explanations, we point to potentially fruitful areas of further research on this issue. With their close alignment to the precepts of reform-based instruction, continued studies employing performance assessments have important contributions to make to the ultimate goal of attaining high levels of achievement for all students.

Acknowledgments

This study was conducted with support from National Science Foundation grant ESI-0196127. The opinions expressed herein are those of the authors and not necessarily those of the funding agency. The authors wish to acknowledge STEP-uP Principal Investigators Linda Mooney and Paul Kuerbis as well as STEP-uP Assessment Coordinator Cynthia Pechacek for sharing their time, expertise, and materials in support of this study.

References

Amaral, O.M., Garrison, L., & Klentschy, M. (2002). Helping English learners increase achievement through inquiry-based science instruction. *Bilingual Research Journal 26*(2), 213-239.

American Association for the Advancement of Science. (1989). *Science for all Americans.* New York: Oxford University Press.

Baker, E. (1997). Model-based performance assessment. *Theory Into Practice, 36*(4), 247-254.

Binder, A. (1984). Restrictions on statistics imposed by method of measurement: Some reality, some myth. *Journal of Criminal Justice*, *12,* 467-481.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2^{nd}Ed.). Hillsdale, NJ: Erlbaum.

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

Geier, R., Blumenfeld, P. C., Marx, R. W., Krajcik, J. S., Fishman, B., Soloway, E., & Clay-Chambers, J. (2008). Standardized test outcomes for students engaged in inquiry-based science curricula in the context of urban reform. *Journal of Research in Science Teaching, 45*(8), 922-939.

Hoover, H. D. (1984). The most appropriate scores for measuring educational development in elementary schools: GE's. *Educational Measurement: Issue and Practice, 3*(4), 8-14.

Jaccard, J., & Wan, C. K. (1996). *LISREL approaches to interaction effects in multiple regression.* Thousand Oaks, CA: Sage Publications.

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

Kim, J. O. (1975). Multivariate analysis of ordinal variables. *American Journal of Sociology*, *81,* 261-298.

Klein, S. P., Jovanovic, J., Stecher, B. M., McCaffrey, D., Shavelson, R. J., Haertel, E., Solano-Flores, G., & Comfort, K. (1997). Gender and racial/ethnic differences on performance assessments in science. *Educational Evaluation and Policy Analysis, 19*(2), 83-87.

Kuerbis, P. J., & Mooney, L. B. (2008). Using assessment design as a model of professional development. In J. Coffey, R. Douglas, & C. Stearns (Eds.), *Assessing science learning: Perspectives from research and practice*, (pp. 409-426). Arlington, VA: National Science Teachers Association Press.

Labovitz, S. (1967). Some observations on measurement and statistics. *Social Forces*, *46,* 151-160.

Labovitz, S. (1970). The assignment of numbers to rank order categories. *American Sociological Review*, *35*(3), 515-524.

Lawrenz, F., Huffman, D., & Welch, W. (2001). The science achievement of various subgroups on alternative assessment formats. *Science Education, 85*(3), 279-290.

Lee, O., Buxton, C., Lewis, S., & LeRoy, K. (2006). Science inquiry and student diversity: Enhanced abilities and continuing difficulties after an instructional intervention. *Journal of Research in Science Teaching, 43*(7), 607-636.

Lee, O., Deaktor, R. A., Hart, J. E., Cuevas, P., & Enders, C. (2005). An instructional intervention's impact on the science and literacy achievement of culturally and linguistically diverse elementary students. *Journal of Research in Science Teaching, 42*(8), 857-887.

Lynch, S., Kuipers, J., Pyke, C., & Szesze, M. (2005). Examining the effects of a highly rated science curriculum unit on diverse students: Results from a planning grant. *Journal of Research in Science Teaching, 42*(8), 912-946.

Martiniello, M. (2008). Language and the performance of English-language learners in math word problems. *Harvard Educational Review, 78*(2), 333-368.

Messick, S. (1996). Validity of performance assessments. In G. W. Phillips (Ed.), *Technical issues in large-scale performance assessment* (pp. 1-18). Washington, DC: US Government Printing Office, Report No. NCES-96-802. Available from ERIC Document Reproduction Service, no. ED 399 300.

National Research Council. (1996). *National science education standards.* National Committee on Science Education Standards and Assessment. Washington, DC: National Academy Press.

National Research Council. (2000). *Inquiry and the national science education standards: A guide for teaching and learning*. Washington, DC: National Academy Press.

National Research Council. (2001). *Classroom assessment and the National Science Education Standards.* Committee on Classroom Assessment and the *National Science Education Standards*. J. Myron Atkin, Paul Black, and Janet Coffey (Eds.). Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

National Research Council. (2005). *How students learn: History, mathematics, and science in the classroom.* Committee on *How People Learn*, A Targeted Report for Teachers, M.S. Donovan and J.D. Bransford, Editors. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Science Resources Center. (1991). *Microworlds teacher's guide.* Burlington, NC: Carolina Biological Supply.

National Science Resources Center. (1994). *Food chemistry teacher's guide.* Burlington, NC: Carolina Biological Supply.

National Science Resources Center. (1996). *Ecosystem's teacher's guide.* Burlington, NC: Carolina Biological Supply.

No Child Left Behind Act of 2001, Public Law No. 107-110, 115 Stat. 1425 (2002).

Pine, J., Aschbacher, P., Roth, E., Jones, M., McPhee, C., Martin, C., Phelps, S., Kyle, T., & Foley, B. (2006). Fifth graders' science inquiry abilities: A comparative study of students in hands-on and textbook curricula. *Journal of Research in Science Teaching, 43*(5), 476-484.

Shavelson, R. J., Baxter, G.P., & Pine, J. (1991). Performance assessment in science. *Applied Measurement in Education, 4*(4), 347-362.

Shaw, J. M. (1997). Threats to the validity of science performance assessments for English language learners. *Journal of Research in Science Teaching, 34*(7), 721-743.

Shaw, J. M. (2009). Science performance assessment and English learners: An exploratory study. *Electronic Journal of Literacy Through Science, 8*(3).

Stecher, B. M. (1995, April). *The cost of performance assessment in science: The RAND perspective.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco.

Stecher, B. M., & Klein, S. P. (1997). The cost of science performance assessments in large-scale testing programs. *Educational Evaluation and Policy Analysis, 19*(1), 1-14.

STEP-uP (Science Teacher Enhancement Program unifying the Pikes Peak Region). (2003). *Embedded assessment package for ecosystems.* Colorado Springs, CO: Author.

STEP-uP (Science Teacher Enhancement Program unifying the Pikes Peak Region). (2004). *Embedded assessment package for food chemistry.* Colorado Springs, CO: Author.

STEP-uP (Science Teacher Enhancement Program unifying the Pikes Peak Region). (2005). *Embedded assessment package for microworlds.* Colorado Springs, CO: Author.

U.S. Department of Education, Office of Elementary and Secondary Education, *No Child Left Behind: A Desktop Reference*, Washington, D. C., 2002.

Zumbo, B. D., & Zimmerman, D. W. (1993). Is the selection of statistical methods governed by level of measurement? *Canadian Psychology*, *34*, 390-399

Developing an Understanding of Inquiry by Teachers and Graduate Student Scientists through a Collaborative Professional Development Program

Stephen J. Van Hook
The Pennsylvania State University

Tracy L. Huziak-Clark
Bowling Green State University

Julie Nurnberger-Haag
Michigan State University

Lena Ballone-Duran
Bowling Green State University

Abstract

The PRISM program partnered K-12 teachers with science and mathematics graduate students who served as Scientists or Mathematicians-in-Residence in the teachers' classrooms. The teachers and graduate students participated in a Summer Inquiry Institute during which they learned about inquiry-based instruction, and then collaborated to develop and co-teach content-rich, inquiry-based instruction in the teachers' classrooms for one academic year. In the first three years of the program, 27 teachers and 18 graduate students participated. The research study examined how participation in PRISM influenced the teachers' and graduate students' conceptions of inquiry, explored what they learned about inquiry by implementing inquiry together in the classroom, and studied the role that their collaboration played in the development of their conceptions of inquiry. Conceptions and use of inquiry were examined through surveys, online journals, interviews and classroom observations. The results indicate that the teachers and graduate students deepened and expanded their understanding of inquiry. Particular themes emerged related to what the teachers and graduate students learned about inquiry through the act of teaching via inquiry were (a) inquiry engages students' minds not just their hands, (b) discussion is essential for student learning, and (c) teachers need to help develop a classroom culture conducive to inquiry in order for students to be successful with inquiry-based learning. This research indicates that teacher-scientist/mathematician partnerships can be beneficial to both parties when structured within a long-term professional development program that immerses the participants in the inquiry process and provides ongoing support.

Correspondence should be addressed to Stephen J. Van Hook at sjv11@psu.edu, Tracy L. Huziak-Clark at thuziak@bgsu.edu, Julie Nurnberger-Haag at nurnber3@msu.edu, or Lena Ballone-Duran at ballone@bgsu.edu

Introduction

Inquiry as an instructional approach has been a significant component of recent science education reform efforts (American Association for the Advancement of Science [AAAS], 1993; Barrow, 2006; National Research Council [NRC], 1996a, 2000; Rutherford & Ahlgren, 1990). The National Science Education Standards (NSES) (NRC, 1996a; p. 22) define inquiry as:

> A multifaceted activity where students: make observations; pose questions; research in textbooks and other reference materials what is already known; plan and implement investigations; use evidence to explain questions; use tools to gather, collect, and interpret data; propose answers, questions, and predications; and communicate findings.

Subsequent research studies have endeavored to further define inquiry (Anderson, 2002; Crawford, 2000; Newman et al., 2004). Because there are multiple ways to encourage scientific inquiry in the classroom (Bybee, 2000; Martin-Hansen, 2002; Tafoya, Sunal, & Knecht, 1980), inquiry is best represented as a continuum of approaches that employ aspects of inquiry in the NSES definition (Brown, Abell, Demir, & Schmidt, 2006; Eick, Meadows, & Balkcom, 2005; Furtak, 2006; Lee, Buxton, Lewis, & LeRoy, 2006; NRC, 2000).

While science teachers are often aware that inquiry is an approach they should be using in the classroom, individuals' conceptions of inquiry can significantly differ. This is important to consider because teachers' views of inquiry affect their use of inquiry (Wallace & Kang, 2004). For example, a teacher who believes that students are engaged in inquiry when doing a hands-on "cookbook" laboratory may not realize that inquiry can be much more than this. On the other hand, teachers who believe that inquiry only occurs when every aspect of the NSES definition is included in the lesson believe that students are doing inquiry only in a completely open-ended lesson. This leads many teachers to conclude that inquiry is too difficult to do and thus do not attempt inquiry at all (Brown, Abell, Demir, & Schmidt, 2006; Wee, Shepardson, Fast, & Harbor, 2007). Therefore, Keys & Bryan (2001) posit that, "multiple modes of inquiry teaching and learning will invite teachers to engage in participating in inquiry in ways that match their own beliefs and teaching styles" (p. 632), a view that is supported by Blanchard, Southerland, & Granger (2009). In addition, the best choice of inquiry method can depend on many variables, including goals of the lesson, student experience with inquiry, classroom context, and school resources (Settlage, 2007; Songer, Lee, & McDonald, 2003).

Collaboration is important to this process of helping science teachers better understand inquiry and develop their ability to use a range of inquiry approaches in the classroom. In a survey of the research on inquiry teaching, Anderson (2002) emphasized the need for collaboration: "Collaboration is integral not only to the technical dimension of reform endeavors, but to the cultural dimension. … Collaboration is a powerful stimulus for the reflection which is fundamental to changing beliefs, values and understandings." (p. 9). Collaboration is also important for sustaining change. In a recent study of teachers during and after a summer inquiry workshop, Wee et al. (2007) found that teachers implemented less inquiry during the academic year than they had planned during the summer workshop. Wee et al. (2007) concluded that support during implementation of inquiry is critical.

Many reform documents have called for the specific collaboration of scientists and mathematicians with K-12 teachers to improve K-12 education (National Commission on Mathematics and Science Teaching for the 21st Century, 2000; NRC, 1996b). Professional development programs have found that pairing a scientist with an educator can create effective facilitation teams for leading K-12 teacher professional development workshops (Czerniak, Beltyukova, Struble, Haney, & Lumpe, 2005; Duran & Ballone Duran, 2005) and can also influence the college-level teaching of the scientists (Ballone, Czerniak, & Haney, 2005). Other researchers have partnered scientists and teachers in the teachers' classrooms to improve inquiry teaching (Caton, Brewer, & Brown, 2000). Scientists in this context, however, usually do not have a background in inquiry-based teaching and often use instructional practices that do not model inquiry (Schuster & Carlsen, 2009). Thus, an effective collaboration between scientists and educators needs to develop both groups' understanding of inquiry in science teaching as well as an awareness of the differences in professional cultures between these two groups (Drayton & Falk, 2006; Nurnberger-Haag, Huziak-Clark, Van Hook, & Ballone Duran, 2008).

The PRISM program, a National Science Foundation (NSF) GK-12 program, fostered year-long collaborations between graduate student scientists and mathematicians and grade 4-12 classroom teachers to support both groups in developing a conception of the entire inquiry continuum, help them learn how to implement effective inquiry strategies in K-12 and university classrooms, and to provide continual support during implementation of inquiry in the teachers' classrooms. The GK-12 program was created to bring together the content expertise of mathematics/science graduate students and teaching experience of K-12 teachers in order to improve the content knowledge of teachers, the communication skills of graduate students, and the science teaching abilities of both groups. To accomplish these goals, PRISM facilitated year-long partnerships between K-12 teachers and graduate students in the role of Scientists or Mathematicians In Residence (hereafter, scientists) in the teachers' classrooms. The scientist-teacher teams participated in an intensive Summer Inquiry Institute and then co-planned and co-taught science and/or mathematics using inquiry for an entire school year (Huziak-Clark, Van Hook, Nurnberger-Haag, & Ballone Duran, 2007; Nurnberger-Haag et al., 2008). The PRISM program and how its participants' understanding of inquiry developed will be described in detail.

Review of Literature

Defining Inquiry Along a Continuum

A continuum of inquiry teaching approaches appears in *Inquiry in the National Science Education Standards* (NRC, 2000), describing varying levels of student self-direction and teacher direction (p. 29). For the purposes of this research, we constructed an expanded inquiry continuum (Figure 1) to frame this study and present to the participants as part of the Summer Inquiry Institute described in the next section. This inquiry continuum is grounded in the literature of known inquiry practices and shows inquiry in increasing complexity as students have increasing control. The continuum begins with Tafoya, Sunal, & Knechts' (1980) *confirmational inquiry*, defined as an activity assigned to the students for which the main task is to prove what is already known or could be inferred by the students. It continues with *structured inquiry*, for which the teacher provides students with questions, procedures, and information but not the correct or known answer. Next is *guided inquiry*, where students are given a question to answer, but they may or may not have the procedure, or the procedure may be developed as a class (Furtak, 2006). Martin-Hansen (2002) adds two types of inquiry to Tafoya et al.'s descriptions beyond guided inquiry: coupled and open. *Coupled inquiry*, a type of inquiry not given in the continuum of NRC (2000), is a combination of guided inquiry and open-ended inquiry: students begin with a question, investigate the issue, share results and then students engage in open-ended inquiry based on discussions or personal interest. *Open-ended* inquiry is one in which students develop a research question and the procedure or method by which they will answer it, including a data collection plan. The inquiry continuum combines levels of student and teacher participation with inquiry strategies to demonstrate a progression from teacher-centered to student-centered inquiry.

Confirmation → Structured → Guided → Coupled → Open-Ended

Figure 1. Definitions of scientific inquiry of Tafoya et al. (1980) and Martin-Hansen (2002) compiled into one inquiry continuum.

Effect of Inquiry on Student Learning

Several meta-analyses have examined the effect of inquiry-based teaching approaches on student learning (Shymansky, Hedges, & Woodworth, 1990; Shymansky, Kyle, & Alport, 1983; Wise & Okey, 1983). Unfortunately, each study in the analyses used different definitions of inquiry and few studied how inquiry was actually implemented with students. However, despite these difficulties in interpreting the literature, the meta-analyses indicate that research supports inquiry as a pedagogical approach. For example, Shymansky et al. (1983, 1990) found an improvement in achievement, attitude and process skills due to inquiry-based teaching. Recent studies have shown that using inquiry-based teaching methods enhanced 3rd and 4th grade students' abilities with some inquiry tasks (Lee, Buxton, Lewis, & LeRoy, 2006), content achievement and engagement with 8th graders (Lynch, Kuipers, Pyke, & Szesze, 2005), and increased science understanding and inquiry skills for 7th and 8th graders (Geier et al., 2008). In addition, Dean & Kuhn (2006) showed that discovery learning has long-term benefits over direct instruction. While the literature is not uniformly supportive of inquiry teaching (Furtak, 2006), Anderson (2002, p.4) summarizes: “In general, research shows that inquiry produces positive results.”

Anderson (2007) argues that “Inquiry learning is foundational and essential for a first-rate education” (p. 821), but then warns that not all teaching thought to be inquiry-based results in inquiry learning. For example, Pine et al. (2006) found that while inquiry-based teaching improved student learning, simply adopting inquiry-based curricula was not sufficient, because few of the teachers “went beyond a recipe-like approach to the curricula” (p. 481). Thus using new curricula designed for inquiry still does not guarantee that students will participate in actual inquiry-based activities. Four years earlier, Anderson (2002) noted this same problem, indicating that although the teachers had resources designed to help them teach via inquiry in the form of NSF supported curriculum materials, “Generally speaking, however, the materials were not being used in a manner consistent with this philosophy” (p. 7). Thus, while inquiry can enhance student learning, adopting an inquiry-based curriculum is not sufficient to ensure inquiry in the classroom.

In order to teach effectively using inquiry a teacher must be willing and able to move flexibly between many different roles. Crawford (2000) identified ten such roles that a teacher may adopt as they facilitate inquiry in a classroom: motivator, diagnostician, guide, innovator, experimenter, researcher, modeler, mentor, collaborator, and learner (p. 931-2). In order to fulfill each of these roles a teacher must be comfortable with a variety of teaching strategies (e.g. cooperative learning, discovery learning, problem-based learning) in addition to possessing strong content knowledge.

Barriers to Inquiry

There are a variety of difficulties that educators face when beginning to use inquiry to teach science. Welch, Klopfer, Aikenhead, and Robinson (1981) suggest some common barriers for teachers, the most common being perceived difficulty, especially with classroom management. Many educators did not learn science or mathematics through inquiry methods themselves, so they may feel ill-prepared to teach with inquiry methods (NRC, 2000; Weiss, Pasley, Smith, Banilower, & Heck, 2003). If teachers attempt inquiry, they may see students’ initial confusion and lose confidence in inquiry’s effectiveness. In addition, Brown and Melear (2006) suggest that when teachers learn science via an open-inquiry process, they “often experience a loss of confidence in their science knowledge” (p. 954). This idea can lead to further reasons that teachers avoid or reject inquiry as a teaching method. Finally, there are political and cultural dimensions that must be addressed so that teachers can be supported in their quest for more inquiry based teaching practices (Anderson, 2002; Anderson, 2007). Pressure from administrators or parents who are not familiar or confident with inquiry may lead to a teacher’s decision not to practice it. Teachers may also feel like they have to teach only factual knowledge in order for their students to succeed in standardized proficiency tests.

Statement of the Problem

The NSES (NRC, 1996a) state several standards for effective science teaching. These standards focus on inquiry in the following statements: "teachers of science plan an inquiry-based science program for their students" (p. 30), "focus and support inquiries while interacting with students, orchestrate discourse among students about scientific ideas, [… and] encourage and model the skills of scientific inquiry, as well as the curiosity, openness to new ideas and data, and skepticism that characterize science" (p. 32). Based on the review of literature presented, these standards are difficult and teachers are faced with many barriers to accomplishing these tasks. The PRISM program focused on addressing ways that teachers and scientists, learning, planning, and teaching together could effectively overcome the barriers to implementing inquiry-based teaching methods.

Consequently, we asked three main research questions about the effect the PRISM Summer Inquiry Institute and academic year professional development had on the teachers' and scientists' ideas about inquiry:

1. How did participation in PRISM influence the teachers' and scientists' conceptions of inquiry?

2. What did the teachers and scientists learn about inquiry by implementing inquiry together in the classroom during the academic year?

3. What role did the collaboration between scientist and teacher play in the development of the teachers' and scientists' conceptions of inquiry?

Description of PRISM Program

The PRISM program partnered science and mathematics graduate students (scientists) and K-12 teachers. Over a period of three years a total of 27 teachers and 18 scientists participated in the program. The scientists were graduate students in biology, chemistry, geology, mathematics or physics. The teacher participants had at least three years experience teaching in grades four through twelve. The scientists were recruited for the program from incoming graduate students and current graduate students recommended by faculty as having an interest in education. The teachers and scientists participated together in a Summer Inquiry Institute and then co-taught in the teacher's classroom for one academic year. The scientists participated for up to two years. In their first year they were partnered with an upper elementary teacher, while in their second year they were usually partnered with a content specific teacher in grades 7-12. The phases of the PRISM program are highlighted in Figure 2. The design components of the PRISM professional development model were purposefully selected and planned based upon research on effective professional development and self-efficacy beliefs from many studies and reports (Fullan, 1982; Haney & Lumpe, 1995; Loucks-Horsley, Hewson, Love, & Stiles, 1998).

Time Frame	Activities
Phase I	**Summer-** Scientists and teachers participate in learning about inquiry, teaching strategies (e.g., 5E model), participate in inquiries along the continuum, and learn from past PRISM teachers and scientists about co-planning and teaching. Begin to plan for the academic year.
Phase II	**End of Summer-** Co-plan and develop inquiry-based activities that meet the state and local standards.
Phase III	**Academic Year-** Continue to co-plan and co-teach inquiry activities. Monthly professional development meetings to discuss barriers and concerns. Additional support. Academic year observations by external observers.

Figure 2. PRISM program description of activities

Summer Inquiry Institute

Using the inquiry research and guidelines described above, the PRISM faculty designed a five-week Summer Inquiry Institute where science and mathematics graduate students and classroom teachers worked together to learn how to design and implement quality inquiry lessons. Components of the Institute were facilitated by Education faculty, Arts & Science faculty, and PRISM participant teacher-scientist teams from previous years of the program. These role models, describing and modeling inquiry experiences, were key to participant understanding of the challenges and rewards of using inquiry in the classroom (Huziak-Clark et al., 2007).

Scientists and teachers participated in a range of inquiry activities from across the Inquiry Continuum (e.g. Exploratorium foam activity (Exploratorium, 2008), open-ended stream table investigations). The participants' experiences from the activities and the observations they made about the range of inquiry teaching approaches from the different activities fed discussions about the Inquiry Continuum (Figure 1) and effective inquiry-based teaching. During the first three weeks of the Institute, scientists and teachers participated in professional development in which they experienced inquiry-based educational practices such as cooperative learning and contextual teaching and learning (Johnson & Johnson, 1994; Kagan, 1992; Tafoya et al, 1980); learning theory, including learning styles and multiple intelligences (Barba, 1995; Donovan, Bransford, & Pellegrino, 1999); and alignment of curricula with national and state mathematics and science education standards (National Council of Teachers of Mathematics, 2000; National Research Council, 1996a). A learning cycle approach, particularly the 5E model (Bybee, 1997), was emphasized as a useful and effective structure on which to build inquiry lessons (Abraham, 1997). The basic learning cycle employs an exploration phase, an explanation phase, and an extension (or application) phase in this specific sequence (Abraham, 1998) and the 5E model adds an engagement phase at the beginning and an evaluation phase at the end (Bybee, 1997). Critical to the model is that the students explore the topic and then develop the scientific explanation, rather than the traditional approach of providing the explanation and then having the students verify the explanation

(Abraham, 1998). All the lessons and workshops in which participants engaged during the Institute modeled this 5E learning cycle.

In the planning and development phase (the last two weeks of the Institute), the teams applied what they learned about inquiry while working together to develop inquiry-based modules or effectively adapt existing high-quality modules consistent with the recommendations found in the *Ohio Academic Content Standards for Mathematics and Science (Ohio Department of Education, 2003).* In addition, teams peer-taught one of their lessons to the group for feedback and reflection. Peers and education faculty provided feedback in order to improve the level of inquiry of the lessons before they were implemented with students.

Academic Year Professional Development

The lessons designed during the Summer Inquiry Institute were implemented by the teachers and scientists during the academic year in the teachers' classrooms. In addition, the teams continued to design and implement content-rich inquiry-based lessons throughout the academic year. The teachers and scientists reflected on their experiences with inquiry in online journals and during project meetings in which they could interact with other teachers and scientists. One of these project meetings was a monthly regional professional development event for which participants were assigned methods/tasks to try with their students, complete reflections, and discuss at the next monthly event. Their professional development was further enhanced by presenting inquiry-based lessons they developed to other teachers at several regional science and mathematics education symposia.

Research Design and Instruments

The overall evaluation of the three-year program was extensive and involved a combination of quantitative and qualitative research methods. However, for the purposes of determining how the teacher-scientist pairs learned about inquiry, developed inquiry lesson plans, and implemented these plans, we draw primarily on four main data sources to aid in triangulation of our themes: inquiry methods survey, journal prompts, classroom observations, and focus group interviews.

Inquiry Methods Survey

The Inquiry Methods Survey (see Appendix I) was developed by the authors and is a pre/post survey where participants were asked to define inquiry, describe a sample lesson, and describe possible barriers to teaching via inquiry methods. The authors established content validity by having the survey reviewed by five other science educators familiar with inquiry teaching methods and barriers. Modifications were made based on the reviews of the content specialists. The survey has 12 open-ended response questions to determine the state of a participant's ideas, attitudes, and concerns about teaching by using inquiry-based teaching methods. The teachers and scientists completed the survey at the beginning of the Summer Inquiry Institute and at the final academic year project meeting. Three researchers read and tallied the responses of the survey to determine pre/post changes. We compared our coding of the responses to check for internal consistency and did so with more than 90% agreement.

Journal prompts

Each participant was asked to respond to online journal prompts once or twice a month. Many of these journal prompts directly related to inquiry-based teaching and learning in the classroom. The journal prompts were a combination of open-ended questions to encourage the teachers and scientists to reflect as well to ask specific questions needed for reporting to the funding agency. For example, one journal prompt asked teachers and scientists individually to "describe a lesson that you co-taught this week and describe how you employed inquiry during this lesson." Questions like this were used to probe the participants' concept of inquiry and to assess the types of inquiry activities being implemented by the teams.

Classroom Observations

Classroom observations were conducted by the PRISM faculty and staff with science and education expertise. In addition, three times a year, teams were formally observed by an experienced educator who was not a member of the PRISM team. The observers used the Horizon's Classroom Observation Protocol (Horizon Research, Inc., 2000a), the design of which was funded by the National Science Foundation to "measure the quality of an observed K-12 science or mathematics lesson" (Horizon Research, Inc., 2000b). This protocol focuses the evaluation on four main areas: lesson design, implementation, science/mathematics content, and classroom culture. Information from these observations was used to provide objective feedback to the teams as well as to document the level of inquiry teaching during the academic year. Formal comments on the teams' progress and use of inquiry served as an additional way to document the successes and challenges of teaching via inquiry.

Interviews

The teachers and scientists participated in structured individual or focus group interviews at the end of each academic year (Fontana & Frey, 2000). In focus group interviews, participants were grouped as teachers or scientists separately to allow for honest reflection. The interview (see Appendix II) consisted of 16 structured interview questions and was run by the project internal evaluator or one of the evaluator's graduate assistants. The same protocol was used for both individual and focus group interviews, and in the focus groups the participants were given an opportunity to respond to each question individually. The interviews were tape-recorded, transcribed, and coded.

Data Analyses

The journal prompt responses, interview responses, and Horizon reports for all of the participants were analyzed using a grounded theory perspective (Charmaz, 2000; Erickson, 1986; Glaser & Strauss, 1967). To identify emergent themes and assess the use of reflective thinking within the data, three readers of the research team independently reviewed the journal prompts, interview transcripts and Horizon reports. From iterative readings of the journal prompts and evidence, initial codes were subsumed under broad categories (Erickson, 1986). For example, each of the research members noted several themes throughout the journal prompts, surveys, interviews and classroom observations. These themes included inquiry teaching, collaboration, student achievement, and enthusiasm about teaching science. The focus of this paper is from the theme of inquiry teaching. After discussing these specific themes and the examples that all three agreed on, the group determined “sub themes” or specific codes and specifically defined and agreed to meanings of each code (Glaser & Strauss, 1967). For instance, “inquiry engages minds, not just hands,” was decided on as a sub-theme under teaching with inquiry. The research team agreed that there were a breadth of respondents who reported they recognized the importance of discussions for inquiry teaching, the need to develop a classroom culture for inquiry, and that inquiry engages minds, not just hands. The theme of inquiry is the umbrella for each of these important ideas. Again, the research team revisited the data and recoded with these categories or codes in mind (Erikson, 1986). These categories were used in further iterations of data readings by the researchers, who met to negotiate and clarify the themes and their meanings. Once this was accomplished, data that fit each of the themes were coded with that category and later used to elaborate on findings in this study. The research team agreed that in order to establish “fit” all three readers had to agree that the data met the operational definition. Miles & Huberman (1994), refer to this as “an organized assembly of information that permits conclusion drawing and action taking” (p. 11). By using the grounded perspective the researchers were able to triangulate meaning from multiple sources, (interview, observation, and journal entry) so that we were able to “accurately describe what [we] understood, constructing recognizable reality for the people who have participated in the study” (Maykut & Morehouse, 1994, p. 122). The findings and conclusions, drawn from the categories, will be explained in subsequent sections.

Findings

Several important themes were evident in the survey, interviews, journals and classroom observation. These themes detail the changing notions and, more importantly, the practice of teachers and future Arts & Science faculty and instructors towards inquiry as a way of teaching. These themes were: viewing inquiry as a continuum, realizing that inquiry engages students’ minds not just their hands, discovering the importance of discussions, and needing to develop a classroom culture that supports inquiry. We document each theme with the inquiry survey, teacher and scientist journals, and interviews and classroom observations as appropriate.

Seeing the Inquiry Continuum

The participants formally reflected on what inquiry meant to them several times during the year in the inquiry surveys and in their journals.

Initial conceptions of inquiry. The teachers and scientists commonly used several key phrases during the pre-inquiry survey. For example, the majority of the participants included the idea of hands-on activities as part of their notion of inquiry. Half of the participants included questions as a key component of inquiry.

The following quotations from teachers are typical of responses on the first day of the Summer Inquiry Institute and depict teachers' early notions of inquiry (NB: All names are pseudonyms. T after the name indicates a teacher; S indicates a scientist):

> *Learning through inquiry is student directed. The students determine the topic of study, method of presentation, and the steps needed to complete the project. The role of the teacher is to guide and move the team in the right direction without giving too much information.* (Mike-T)

> *Learning is more student lead and less teacher lead. The students have more control over what they are experimenting with and why.* (Marjorie-T)

> *I would define learning through inquiry as a technique that can be used to help some students. This technique involves the student seeking the answers to questions instead of simply being told the answer.* (Herman-T)

The teachers hold that inquiry is student driven and, upon further discussion, most believed that students have to be involved in individual or group projects instead of a class investigation for it to be considered inquiry. Thus, the teachers mostly began with an understanding of inquiry as primarily open-ended, placing their conception of inquiry at the right end of the Inquiry Continuum (see Figure 1). Their understanding does not reflect the flexibility of inquiry-based teaching methods described in the review of literature. Teachers' use of a strictly open-ended conception of inquiry instruction was itself a barrier to doing inquiry in their classrooms. By helping teachers recognize that inquiry can take many forms across the continuum, we help them bridge the gap between the position on the continuum where most teachers' lessons would be classified and the open-ended region of the inquiry continuum. In this way teachers can see how they can move themselves and their students toward the open-ended portion of the continuum. In this way they find validation at the other points along the continuum knowing that they now have the skills and understanding to facilitate inquiry-learning at multiple degrees and that open-ended inquiry is not always the best method for accomplishing particular learning goals.

The scientists had little to no formal education in pedagogy, so their ideas about teaching math and science stemmed largely from their experiences as students in undergraduate content courses or their K-12 school experiences. The following early scientists' notions of inquiry from the highlight their thinking at the start of the Summer Inquiry Institute:

> *Giving students examples and having them come up with the definitions and methods.* (Jacob-S)

> *Teachers asking questions to spark discussions.* (Ralf-S)

> *By asking questions to discover the individual answer, or solution to a given problem.* (Ruth-S)

Note that the scientists' initial ideas are mainly focused on the teacher, not the students. These ideas about inquiry mostly sit near the left end of the Inquiry Continuum (Figure 1) and are, for the most part, different than those of the teachers.

Conceptions of inquiry one year later. As the quotations above illustrated, the scientist initial views of inquiry tended to reflect the confirmation side of the continuum as being representative of inquiry, while the K-12 teachers often viewed only very open-ended learning as inquiry. In the post inquiry survey, after participating in the Summer Inquiry Institute and collaborating with their partner to implement inquiry in the classroom for the academic year, most of the participants expanded their conceptions of inquiry to cover a larger band of the Inquiry Continuum. As one teacher, Barbara, expressed it during her end-of-year interview,

> *I'm trying to apply what I learned just like the kids apply what they learned, but then I try to go back and reflect on that and see how can I make it more inquiry based, or is this an appropriate level, and what lesson fits in where with the inquiry continuum and all that stuff.* (Barbara-T) (Huziak-Clark et al., 2007)

The following post inquiry survey quotations provide examples of these post notions of inquiry for teachers:

> *Learning through inquiry is when students are responsible for their own learning. It is when a question (either student or teacher generated) is posed to the student and it is up to the student to determine the answer. That answers are then compared to the rest of the class and as a large group the result are discussed to determine the results of the lab. It is this discussion of lab results that drives the student learning*. (Mike-T)

> *Learning through inquiry is a global approach for the students. Inquiry brings together background knowledge, reading, writing, and working together to solve or answer questions one has about science or any other topic.* (Chelsea-T)

One significant change in the teachers' ideas of inquiry was the idea that inquiry is a communal effort by the class, not just by individuals. This is an important change in beliefs that can lead to changes in practice. Teachers' prior beliefs that every student had to do their own project for it to be inquiry could be a daunting notion for even experienced teachers. In addition, inquiry without student-student interactions ignores the critical role that social interactions provide in developing understanding. Being able to view inquiry as something that a class can contribute to and work on collaboratively is an important shift. In addition, the collaborative nature of this instruction reinforces that scientists often work in teams.

The scientists' understanding of inquiry also changed over time through working with the teachers and by participating in the Summer Inquiry Institute. For example, Taylor provided the following conception of inquiry in the post inquiry survey:

> *Learning through inquiry involves the students discovering concepts on their own through minds-on and hands-on experiences. Inquiry does not necessarily have to be a hands-on experience. The only requirement is that the student has ownership in his/her development.* (Taylor-S)

This idea marks a shift from primarily teacher-centered conceptions of inquiry that the scientists held to a more student-centered conception of inquiry. Like Taylor, Ruth's conception also showed a shift towards thinking about what the students do, not just what the teacher does:

> *Presenting the students with a hands-on science experiment, and asking questions that require them to think deeper about the experiment. Also, having the students explain to each other what they think is happening and why!* (Ruth-S)

In addition, notice how Ruth's conception of inquiry now includes the core idea that inquiry can be a group/class endeavor, one of the shifts noted earlier for the teachers.

The above examples provide a view of the teachers' and scientists' notions of inquiry from the inquiry survey. The teachers and scientists also reflected about inquiry in their journals:

> *Inquiry means three things to me: discussion, critical thinking, and hands-on. The discussion can be taking place between teacher and students or students and students. […] There is no doubt that it would be easier as a teacher to simply tell or lecture the information to the students because it takes less time. However, I do believe students get more out of an inquiry lesson than a lecture lesson. Students feel a sense of accomplishment by "figuring something out for themselves."* (Ralf-S)

> *The cooperative learning strategies have provided my lessons with a more structured and constructivist approach. The 5E learning model reminds me to value student questioning and to build lessons around big ideas. During my science lessons, I am no longer in the front of the classroom. Rather, I am facilitating small groups of young scientists, while posing problems of relevance.* (Tracy-T)

The change here from the initial survey is a more holistic approach to inquiry. More of the teachers and scientists see inquiry as a way of thinking about teaching than just a pedagogical technique. From the quotations shown, one can note the shift to an emphasis on the cognitive aspects of an inquiry-based lesson, the importance of questioning and discussions, and the need for specific skills for students to do inquiry. Each of these themes is expanded upon below.

Inquiry Engages Students' Minds Not Just Their Hands

Although hands-on activities are often a key element in inquiry-based lessons, they are neither always necessary nor sufficient by themselves. A clear theme in the classroom observations, journal prompts, the survey, and interviews was that participants developed an understanding that inquiry is more than just hands-on activities. Before the Summer Inquiry Institute many participants initially believed that having students complete an activity about the content was sufficient for students to learn the content. Through a year of co-planning and co-teaching, the participants recognized that teaching by inquiry requires engaging the students' minds, not just their hands. Below a scientist describes tweaking his partner teacher's existing hands-on lesson to add an inquiry component:

> *We have started our weather unit that [teacher partner] and I worked on this summer. Our first few lessons have been on the nature of air. Each of our lessons on air has begun with a "challenge" to the students. Our first challenge was for [the students] to get air into a submerged, overturned cup without taking it out of the water or turning it over. The only tool they could use to get air in the cup was a smaller cup. At first many of the students thought the task was impossible however they all dove-in with enthusiasm. After a little bit of trial and error they hit upon the solution of holding the smaller cup upside down so air stayed in it until they got it under the larger cup and then turn it over so the air would go into the larger cup. The students then discussed in their groups what this showed about air (i.e. air takes up space). [Teacher partner] said that she had done this lesson previously but that in the past she had demonstrated to the students the solution beforehand. She said that those students had thought the lesson was neat, but that this one really got the students involved and made them think* (Chester-S).

Both the teacher and scientist realized that even though a hands-on component was already present, making the lesson more inquiry-based required getting the students mentally engaged in the content of the activity. Collaboration with the scientist was critical in helping the teacher make this change since the scientist encouraged the teacher to let the students figure out the solutions on their own and then provided support during the lesson. Many of the scientists discussed their efforts of promoting student thinking in their journals. For example, a scientist described his experience with a spaghetti tower construction project early in the school year:

> *The students came to me for assistance when their tower designs did not materialize in the manner they had hoped. Naturally, what the students wanted was for me to tell them what was wrong with their designs and to do X to fix problem Y. Just as naturally this was not something I was going to do. I tried to direct their attention in a critical manner toward their towers, asking them to identify the problem/weakness in their structure and then think about how they might address the problem/weakness. […] My goal was not the success or failure of the tower but to lead the students through the process of analytical thought and problem solving.* (Mark-S)

A month later, the same scientist commented in his journal,

> *Now that the students have a few design activities under their belts (both successful, and perhaps more importantly unsuccessful) their approach is completely different. […] Many students are now looking to identify the cause of problems in their racers and not simply disassembling and rebuilding during the problem-solving phase. There are even a few students whose racers incorporate genuinely surprising and completely functional design elements.* (Mark-S)

Notice that now with these hands-on activities the scientists are focused on the mental processes of the students, and work with the students to develop these processes further. The teachers echo the scientists in their journals and in interviews about their PRISM experiences.

> *[PRISM has] helped fine-tune me too. Now I can do hands-on with the best of them, but it's not always inquiry. And that's the difference. So now I'm learning how to put the inquiry in the whole package. As opposed to just the experiment part.* (Chelsea-T)

> *I am enjoying the inquiry teaching, but am having to hold myself back from giving the answers rather than letting them DISCOVER on their own! It was really neat seeing the students get excited about learning and asking their own questions and finding their own answers.* (Valerie-T)

The teachers are each explaining that they previously did hands-on activities; however, now due to their experiences in the Summer Inquiry Institute and their collaboration with the scientists, they are reevaluating the methods they use, how they present these laboratory experiences, and how these hands-on activities are now just one aspect of a more developed lesson. The teachers realized that just because a lesson involved a lab or a hands-on activity did not always mean that the lesson was inquiry-based. They have begun to realize that, as Songer, Lee, and McDonald (2003) wrote, "Inquiry is more about substance than form" and that it is the "quality of intellectual engagement among students" that is a critical factor in inquiry (p. 514).

Importance of Discussions to Learning Via Inquiry

As the participants began to realize the importance of student thinking in inquiry, they also began to recognize that in order to foster true student learning, they needed to change their planning in order to make more time for discussions and "sense-making" of the content. For example, a major difference between teacher Mike's pre- and post-survey notions of inquiry is his recognition of the importance of discussion in the post-survey, which concludes with, "It is this discussion of lab results that drives the student learning."

A common theme in the journal responses, the interviews and the surveys is the importance of discussion and sharing ideas for students to learn from the inquiry process. Many teachers and scientists began to realize that they could engage their students in inquiry by facilitating a rich discussion about the topic first. Not only did it get the students motivated and interested, it also helped the teacher understand what the students already knew. As one teacher (Elinor) explained in the end-of-year teacher focus group,

> *You always have to begin with the questions. You always have to have a question that you can manage and change, and so I got that now. And the more I can have my kids think about a question and then begin what they want to do with that question. That's the goal.* (Elinor-T)

The critical importance of the class discussion during the explanation phase of the 5E learning cycle was new to many of the participants. In particular, the participants realized that the explanation should draw from the students' observations and ideas in the context of a discussion about the topic. In one of her journal entries, a scientist described how discussions are the difference between doing inquiry and just doing an activity:

> *I believe two classes could do the exact same lab activity, and depending on how it is approached and discussed determines whether it is an inquiry-based lab or just an activity. I think understanding the process of asking questions to lead the students to the information and not just dumping the information on them is the key. […] The more you can get teachers to understand the importance of a good discussion, and get them to feel totally comfortable leading one, the better inquiry-based labs you'll have going on in the classroom.* (Kerstin-S)

Another teacher described how a lesson now flows in her classroom. Notice she now understands the purpose of the explanation phase in the 5E learning cycle. For this teacher, *explain* is no longer synonymous with teacher lecture or exposition. Students and/or teachers can participate in doing the explaining as long as the necessary connections and clarifications are made to tie the lesson together.

> *Students record observations and then do group discussions to talk about their observations. Then the groups make small presentations about their findings. During these discussions a lot of the Explain comes out. Students have questions about why certain things happened so it leads into the explanation quite easily.* (Gertrude-T)

The scientist (Peter) who co-taught with Gertrude described one of these class discussions:

> *Unifying the results of the class and drawing conclusions was very tricky. One group disagreed with the rest of the class in their results, so we watched them demonstrate the experiment. It was great – the class ripped their misconceptions to threads. Once we had a consensus, we could look at some theoretical applications of Bernoulli's Principle.* (Peter-S)

These quotations illustrate how the participants viewed class discussions, including student presentations and peer review, as a vital mechanism for sense-making by the students. They believed that these discussions were critical for clarifying student ideas and correcting their misconceptions. Notice also that the starting point of the explain phase was the students' observations, ideas, and experimental results. The participants drew upon the students' own data to lead to the students to a concept, rather than just telling the students what to think. Teachers discussed how their students had noticed the change:

> *The students in my classes have come to realize that lab is not just something you get through to say, "Ok we finished that now let's move on." They realize there is a valuable lesson to be learned from doing the lab*. (Mike-T)

> *I have found out that the students need to bring concepts to a close. They really get into discussions and want to take things to the next level.* (Kelly-T)

Notice the importance described by both participants that just completing the activity is not enough. They both recognized that students will learn more if they are provided the time to discuss and process the material in groups and as a class. In Kelly's case, her comment at the end of the year showed a major shift in her thinking since Gertrude, the scientist with whom Kelly worked, had set a goal for the year of helping Kelly recognize the importance of using discussions after hands-on activities instead of just moving on to the next lesson. Mike's partnering scientist also promoted expanded post-lab discussions throughout the year, which were emphasized in both of Mike's previous comments. The scientists also gained a greater appreciation for the role of discussions for learning at the university level. Jacob described how he taught a university mathematics course while involved in the PRISM project:

> *[It] has been interesting to see how my approach to teaching has changed. A few weeks ago I covered a section on transforming the graphs of functions. … When I got to class, I was surprised at how I was teaching. Instead of lecturing, I was having a conversation with the class. I would ask the students questions, and then take the students answers and comments and get us to where we needed to be. … As I was teaching, I found myself thinking more about what the students had to say, than what I wanted to say. I was interested in allowing them to lead the discussion and have the class learn from what they had previously seen.* (Jacob-S)

The external reviewers also noted the use of good discussions and questioning that was evident in teams' planning and implementation:

> *The design reflected careful planning and organization and allowed adequate time and structure for sense making for the students. […] The activities led to an in-depth discussion that allowed Julian [the teacher] to recognize the level of understanding of his students. This was prompted by Julian asking higher-order questions about his students' responses to the brainstorming wheel.* (Paula-observer)

> *Through the use of discussions students were able to use mathematical reasoning and justification of ideas. By working collaboratively as a class, students were able to constructively challenge each others' ideas. The activities in the lesson encouraged investigation of questions and providing justification for answers.* (Sheila-observer)

These comments further demonstrate the value the participants felt discussions played in teaching the lessons via inquiry. In addition, these reports support the implementation of the participants' planning into practice.

Developing A Classroom Culture That Supports Inquiry

Finally, many of the teachers and scientists found that their students were prepared to complete "cookbook" laboratory activities, but when it came to implementing more inquiry-based lessons, many skills such as questioning and working in groups needed to be developed and enhanced. In journal posts, teachers discussed the classroom culture required for student success.

> *Children need to have good questioning skills in order to perform inquiry-based lessons. They also have to be able to use higher level thinking, and they need the confidence to know that they are able to do this.* (Melissa-T)

Several of the teachers commented on this culture-building problem as one of the initial barriers to inquiry for them in the past. As one teacher described her situation at the beginning of the year in her journal:

> *Our fourth grade students need a lot of skills that they don't necessarily have to participate in inquiry lessons. The biggest problem I've experienced is with the kids that don't know how to work in cooperative groups. In our school they don't do a lot of that in third grade so when they get to fourth grade it can sometimes be a tough adjustment. Once we learn how to work in groups, it gets a lot easier.* (Kate-T)

Another teacher described similar concerns and discussed in her journal how inquiry actually helps develop these skills, rather than viewing these skills as a prerequisite barrier to beginning to teach via inquiry.

> *[My partner scientist] and I have been working on improving the students' cooperative learning skills and increasing their collaboration. This goal is coming along very nicely. Inquiry-based lessons actually help them develop these skills. On Thursday, we will get a better idea of how well the students can use these skills. The lesson will be at a higher level (less structure) of inquiry. (Increasing the level of inquiry has also been a goal of ours.) They will be required to do more problem-solving and reasoning.* (Carie-T)

Both teachers and scientists noted the initial difficulty students can have with inquiry since it differs from traditional school instruction:

> *When we PRISM fellows come into the classroom with our inquiry-based teaching we put the students out of their element. Both traditionally successful and unsuccessful students can resist inquiry because it's different from the school thing they've been doing for the past 12-13 years. By starting off with easier inquiry lessons and practicing inquiry we can smooth the initial resistance and build up a confidence and comfort with a new technique for learning.* (Mark-S)

Note how Mark is thinking in terms of the Inquiry Continuum (Figure 1) in gradually moving his students from the teacher-directed end of the continuum to more open-ended forms of inquiry.

It was not just the PRISM participants who needed support in changing to an inquiry-based way of learning. Participants noted that students are used to teachers emphasizing the importance of the final answer. Consequently, participants reported that students recognized a change in the classroom culture from the way they were used to learning to a culture of inquiry learning. Students recognized a shift in emphasis from obtaining results or answers to the process of learning. For example, one scientist, Eddie, noted in his journal,

> *[W]e have to spend a fair amount of time in the Explain stage and often have to remind kids to think about what they already know. It might seem strange, but we often have to remind our kids to stop and think as they just want to get to an answer and move on.* (Eddie-S)

Similarly, a teacher described in an interview how he worked with the students to value the process of learning, not just getting the final answer:

> *Kids instead of just writing and listening they're actually having some experiences with trial and error, they can participate. I told my students it was the hardest thing, it'll be hard for you too because your whole life you've done it this way: here's some notes, here's what you're supposed to find, do it. [… The students] were saying, "Man, I had to think so much with this." Absolutely, that's what I want. That's actually another thing I have learned – the kids have been trained to do it this other way so you have to get them used to doing it as well. To not being told exactly what they're supposed to find, but it's okay to mess up and not do things right and to learn from it.* (Julian-T)

Because this shift to inquiry can be a significant culture change for students, many participants discussed the motivational aspects of inquiry, in particular the need to build the students' confidence in their ability to do inquiry. For example, scientist Marius observed that,

> *For the students to succeed in inquiry, I feel that they need to be encouraged enough to believe that they can find answers on their own. They still need guidance, but the confidence in themselves will get them*

> *very far on their own. The more they attempt on their own the more they will get out of the lessons.* (Marius-S)

In addition, students need to feel comfortable discussing ideas. Scientist Ralf suggested that a classroom culture for discussion is needed for inquiry learning to take place.

> *I think the biggest and simplest aspect for our students to succeed in inquiry is having them feel comfortable and confident in discussing their thoughts and opinions in science. Tracy does a great job in creating this atmosphere in the classroom.* (Ralf-S)

Not only is it necessary to facilitate discussions, but, as the quotation above described, it is important for students to feel safe and that they have a voice in the discussion. It is not only the job of the teacher to discuss, but of the whole class community to participate in discourse. The scientists were learning from the teachers how to establish classroom cultures that can support inquiry.

The Horizons Observation protocol explicitly examines classroom culture and the external evaluators noted a classroom culture for inquiry as a strength for many of the teams. For example, Paula said this about one team: “there was a climate of respect for students’ ideas, questions and contributions. Interactions between teachers and students reflected collaborative working relationships.” Dawn stated it in a slightly different way: “There was a climate of respect and the teachers had an awareness of what was going on in the classroom and whether or not more information was needed at any given time. The climate of the lesson provided opportunities for students to brainstorm, make conjectures, and ask questions in a safe environment conducive to learning.”

There are many factors that were important to making inquiry successful in each of the team’s classrooms. As a strategy new to many students, the teachers and scientists needed to facilitate the students’ transition to inquiry-based learning through developing inquiry skills, building students’ confidence, and, perhaps most difficult, changing the mindset of the students – so that ambiguity and making mistakes are acceptable and getting to the “right” answer is not the most important goal of the lesson.

These changes in the teachers’ and scientists’ understanding of inquiry did not occur immediately, or even by the end of the Summer Inquiry Institute. In their journals throughout the year they would share new revelations or understandings about inquiry. This was particularly true for the scientists since inquiry teaching was a newer concept to them than to the teachers. Tina (scientist) reported in February of her first year, “suddenly realizing what PRISM has been trying to teach me all along. I really finally internalized the idea of inquiry!” Another scientist (Mark) described in his journal in May at the end of the year his journey with inquiry:

> *During the summer session I remember being highly skeptical of this 'inquiry thing'. I was afraid we were being peddled the latest and greatest development in science education practice, absolutely guaranteed to make*

> *your students 75% smarter, cut gender/socioeconomic achievement differences, and cure male pattern baldness in the class hamster. … By the end of the summer session I figured that there might be something to this 'inquiry thing' in theory, but I still wasn't sure that we could make it work in the classroom. … [A]s the year progressed we tried many ambitious lessons using inquiry as the vehicle to deliver content. Some were successful, others less so. We became more confident in our ability to use inquiry as a tool to deliver content material and make the lessons succeed with greater frequency. …*
>
> *In the last couple of weeks [teacher partner] and I have been running our pond water investigation with the students. This is the most ambitious open inquiry we have done all year. The students are given a sample of pond water and challenged to design an investigation to learn something about it. For the first couple of days the students really struggled with this inquiry. … After a few days of grinding through this frustration the student have begun picking things up. They're beginning to put the pieces together. You can see the wheels starting to turn and the student's creativity starting to show. The students are starting to get excited about things they are discovering in the pond water.*
>
> *The epiphany? Inquiry isn't about experiments. It's not about leading the students to a bit of content knowledge, nor is it about learning a process. Inquiry is about using thought to fight through a wall of 'not understanding', breaking through and turning around to finding a door was there the whole time.* (Mark-S)

Limitations of Study

The inquiry survey, journals, and end-of-year interviews are data reported by the participants themselves, though the external observers' reports are consistent with the self-report data. Many factors affect changes in a classroom environment and we cannot isolate the effect of just the PRISM program on the teachers. In addition, we cannot completely separate out the effects of the Summer Inquiry Institute and academic year meetings from the effects of the co-teaching collaborations. Finally, the teachers in the study were generally those who seek out professional development opportunities and the graduate student scientists went through a rigorous application process that also probed their interest in K-12 education. Thus, the participants were not a random cross-section of teachers or graduate students. Yet, it is interesting to note that precisely despite being experienced educators who seek to improve their practice, their initial concepts of inquiry were quite limited and for the most part untried with students until the PRISM program.

Conclusions

Participation in the year-long PRISM professional development program led to changes in teachers' and graduate student scientists' notions and attitudes about using inquiry teaching. Based on the research questions investigated, we drew several conclusions.

First, how did participation in PRISM influence the teachers' and scientists' conceptions and use of inquiry? The teachers expanded their understanding of inquiry beyond open inquiry, which they had rarely used, to the full continuum of inquiry approaches. From the teachers' journals, interviews and classroom observations, it is clear that teachers were learning to use inquiry across the continuum and were beginning to implement it effectively in their classrooms. Success in overcoming some of the barriers to inquiry will likely make them more willing to try new inquiry lessons. In the interviews at the end of the year, the teachers stated that they would continue to build on those lessons developed with the graduate student scientist or mathematician. As was noted by the scientists in journals, interviews and in classroom observations, there has been a change in their thinking about inquiry teaching, too. They have expanded their understanding of inquiry from the teacher-centered end of the continuum to the entire continuum and have begun to see its value for teaching, even at the university level.

Second, what did the teachers and scientists learn about inquiry by implementing inquiry together in the classroom during the academic year? The participants developed an appreciation for the importance of both the cognitive and social aspects of inquiry. They recognized that hands-on activities alone do not constitute inquiry; rather, inquiry engages students' minds, whether it is confirmation, structured, guided, coupled, or open-ended inquiry. They learned that discussions are critical to student learning via inquiry. They realized that discussions can be used throughout the learning cycle – to engage the students' thinking at the beginning of the lesson or to allow for sense-making by the class after an activity. Finally, they discovered that inquiry requires a classroom culture that promotes investigation and discussion by making students comfortable to ask questions, express and challenge their ideas, and make mistakes. These changes in the teachers' and scientists' understandings of inquiry did not occur immediately, but over the course of the year, as illustrated, for example, by the teachers Kelly and Mike in learning the importance of discussions, and by scientist Mark's journal of his inquiry journey.

Third, what role did the collaboration between scientist and teacher play in the development in the teachers' and scientists' conceptions of inquiry? The partnership was a critical component of the change in the teachers and scientists. In their journals and interviews, scientists and teachers often talked about the importance of their partner. For the teachers, it provided a partner who would provide continual support and feedback during implementation, an important part of teacher implementation of inquiry as discussed in the literature review. In addition, collaboration with a content expert provided the teachers confidence to attempt more open forms of inquiry, as illustrated by the example of Chester and the air lessons described earlier in Findings. Partnership with classroom teachers provided the scientists an opportunity to explore inquiry teaching in an environment where they had support from experienced educators who could establish classroom cultures appropriate for inquiry. Also, the teacher and scientists were able to identify areas in which they could improve and then work together to effect change. The example of Kerstin and her success in helping Kelly learn to use discussions highlights the value of these collaborations.

Implications

Evidence from this study suggests that the collaborative opportunities afforded by the PRISM program expanded and clarified the teachers' and scientists' notions about inquiry. The following implications can be drawn from this study:

> *Extensive and long-term scientist/mathematician-teacher collaborations facilitated by professional development in inquiry-based teaching can be an effective way to change conceptions about inquiry and to promote inquiry-based teaching in K-12 classrooms.*

The Scientist/Mathematician-in-Residence and K-12 teacher partnership was a critical component of the PRISM professional development model. The partnership allowed each partner to employ inquiry with the constant support and feedback of another professional. Each partner brought to the team complementary strengths that were critical in overcoming barriers to implementing inquiry. In addition, the collaboration held both partners accountable for implementing inquiry, thus ensuring that teachers and scientists had the opportunity to learn by using inquiry. Finally, collaboration allowed joint reflection and continual feedback for improving both partners' pedagogical skills.

> *Positively influencing teacher and scientist notions about inquiry-based teaching requires time and experience using inquiry, and is aided by a support structure that encourages the use of inquiry and reflection about the use of inquiry.*

The Summer Inquiry Institute provided the teachers and scientists with experiences using a variety of inquiry approaches and provided a foundation for the academic year. It was the intensive, year-long teacher-scientist collaboration, however, that helped them experience inquiry in their own classroom and see its effect on the students. This success in the classroom reinforced what they learned in the Institute and addressed their doubts about inquiry. The participants described changes in their understanding of inquiry due to their experiences in the classroom, even late in the academic year. Having a collaborator in planning and teaching, as well as reporting regularly to the project staff through journals and in project meetings, provided the participants a support structure that encouraged using inquiry and reflecting on their experiences.

Directions for Future Research

This study suggests several areas of future research about scientist-teacher collaborations to enhance inquiry-based teaching. First, do the teachers continue to use the inquiry lessons they have developed with the scientists and grow in their use of the inquiry approach, or do they use less inquiry over time without the continued support of the scientists and the PRISM program? Second, how do the scientists who become higher education faculty employ inquiry in their higher education classes? Are they able to translate their understanding of inquiry teaching from a K-12 setting to a higher education setting? Third, in what ways do the scientists collaborate in formal or informal ways with K-12 teachers later in their career? Fourth, how could this model be employed with Ph.D. scientists instead of graduate students in a realistic manner? For example, could a K-12 collaboration sabbatical program be created to facilitate higher education faculty to serve as scientists in residence at a K-12 school. Finally, what is the effect on the K-12 students' attitudes towards science and a career in science from having a scientist in residence in their classroom for a year?

A vision of science teaching and learning promoted by many reform documents calls for science and mathematics classrooms to become active and inquiry-based environments. The lessons learned by PRISM provide an insight into a model of professional development that supports the participants in pushing through the barriers to teaching using inquiry. The experiences explained here are also unique because there is a true collaboration between classroom teachers and graduate student scientist/mathematicians, with both parties developing knowledge about inquiry-based teaching and learning. If we expect this type of teaching to occur in future science classrooms, it is important that as researchers we recognize and document the time, effort, professional development, and university support necessary to aid classroom teachers and future university science and mathematics faculty in the art of inquiry-based teaching and learning.

Acknowledgments

Funding for the PRISM program came from National Science Foundation GK-12 Grant #DGE-0231853. The authors also wish to thank the Center of Excellence in Science and Mathematics Education (COSMOS) at Bowling Green State University, the participating K-12 teachers and graduate student fellows, and Bowling Green State University faculty and staff for their support of the PRISM program.

References

Abraham, M. (1997). The learning cycle approach to science instruction. *Research Matters-To The Science Teacher*, No. 9701. The National Association for Research in Science Teaching.

Abraham, M. (1998). The learning cycle approach as a strategy for instruction in science. In B. J. Fraser and K. G. Tobin (Eds.) *International Handbook of Science Education* (pp. 513-524). Hingham, MA: Kluwer Academic Publishers.

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

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

Anderson, R. D. (2007). Inquiry as an organizing theme for science curricula. In Sandra K. Abell and Norman G. Lederman (Eds.) *Handbook of Research on Science Education* (pp. 807-830). Mahwah, NJ: Lawrence Erlbaum Associates.

Ballone, L., Czerniak, C. M., & Haney, J. (2005). A descriptive study of the effects of a LSC project on scientists' teaching practices and beliefs. *Journal of Science Teacher Education*, 16(2), 159-184.

Barba, R. (1995). *Science in the multicultural classroom.* Needham Heights, MA: Allyn & Bacon.

Barrow, L. H. (2006). A brief history of inquiry: From Dewey to standards. *Journal of Science Teacher Education*, 17(3), 265-278.

Blanchard, M. R., Southerland, S. A., & Granger, E. M. (2009). No silver bullet for inquiry: Making sense of teacher change following an inquiry-based research experience for teachers. *Science Education*, 93(2), 322-360.

Brown, P. L., Abell, S. K., Demit, A., & Schmidt, F. J. (2006). College science teachers' views of classroom inquiry. *Science Education*, 90(5), 784-902.

Brown, S., & Melear, C. (2006). Investigation of secondary science teachers' beliefs and practices after authentic inquiry-based experiences. *Journal of Research in Science Teaching*, 43(9), 938-962.

Bybee, R. (1997). *Achieving Scientific Literacy: From purposes to practices.* Portsmouth, NH: Heinemann Publications.

Bybee, R. (2000). Teaching science as inquiry. In J. Minstrell & E. H. van Zee (Eds.), *Inquiring into inquiry learning and teaching in science* (pp. 20-46). New York: American Association for the Advancement of Science.

Caton, E., Brewer, C., & Brown, F. (2000). Building teacher-scientist partnerships: Teaching about energy through inquiry. *School Science and Mathematics*, 100(1), 7-15.

Charmaz, K. (2000). Grounded theory. Objectivist and constructivist methods. In N.K. Denzin, & Y. Lincoln, (Eds.) *Handbook of Qualitative Research, 2nd Ed.* London: Sage Publications, Inc.

Crawford, B. (2000). Embracing the essence of inquiry: New roles for science teachers. *Journal of Research in Science Teaching,* 37(9), 916-937.

Czerniak, C. M., Beltyukova, S., Struble, J., Haney, J. J., & Lumpe, A. T. (2005). Do you see what I see? The relationship between a professional development model and student achievement. In R. E. Yager (Ed.), *Exemplary science in grades 5-8: Standards-based success stories*. Arlington, VA: NSTA Press.

Dean, D., & Kuhn, D. (2007). Direct instruction vs. discovery: The long view. *Science Education,* 91(3), 384-397.

Donovan, M., Bransford, J., & Pellegrino, J. (Eds.). (1999). *How People Learn: Bridging Research and Practice.* Washington, DC: National Academy Press.

Drayton, B., & Falk, J. (2006). Dimensions that shape teacher-scientist collaborations for teacher enhancement. *Science Education*, 90(4), 734-761.

Duran, E., & Ballone Duran, L. (2005). A model staff development program and its impact on early childhood teachers' self-efficacy. *Journal of Elementary Science Education,* 17(2), 1-12.

Eick, C., Meadows, L., & Balkcom, R. (2005). Breaking into Inquiry. *The Science Teacher,* 72(7), 49-53.

Exploratorium. Foam activity: Three kinds of hands-on science. Retrieved July 21, 2008, from http://www.exploratorium.edu/IFI/activities/foam/foam1.html.

Erickson, F. (1986). Qualitative methods in research on teaching. In M.C. Wittrock (Eds.). *Handbook of Research on Teaching*, (3rd ed.). London: Macmillan Publishing Company.

Fontana, A., & Frey, J. (2000). The interview: From structured questions to negotiated text. In N.K. Denzin, & Y. Lincoln, (Eds.) *Handbook of Qualitative Research, 2nd Ed.* London: Sage Publications, Inc.

Fullan, M. (1982). *The Meaning of Educational Change*. New York: Teachers College Press.

Furtak, E. M. (2006). The problem with answers: An exploration of guided scientific inquiry teaching. *Science Education*, 90(3), 453-467.

Geier, R., Blumenfeld, P. C., Marx, R. W. Krajcik, J. S., Fishman, B., Soloway, E., & Clay-Chambers, J. (2008). Standardized test outcomes for students engaged in inquiry-based science curricula in the context of urban reform. *Journal of Research in Science Teaching*, 45(8), 922-939.

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research.* Aldine Publishing Company: NY.

Haney, J., & Lumpe, A. (1995). A teacher professional development framework guided by reform policies, teachers' needs, and research. *Journal of Science Teacher Education,* 6(4), 187-196.

Horizon Research, Inc. (2000a). Classroom Observation Protocol. Retrieved July 21, 2008, from http://www.horizon-research.com/instruments/clas/cop.pdf.

Horizon Research, Inc. (2000b). Retrieved July 21, 2008, from http://horizon-research.com/insidetheclassroom/instruments/obs.php.

Huziak-Clark, T., Van Hook, S. J., Nurnberger-Haag, J., & Ballone Duran, L. (2007). Using inquiry to improve pedagogy through K-12/university partnerships. *School Science & Mathematics*, 107(8), 311-324.

Johnson, D., & Johnson, R. (1994). *Learning together and alone: Cooperation, competition, and individualization*. Englewood Cliffs, NJ: Prentice Hall.

Kagan, S. (1992). *Cooperative learning: Resource for teachers*. Riverside, CA: University of California.

Keys, C. W., & Bryan, L. A. (2001). Co-constructing inquiry-based science with teachers: Essential research for lasting reform. *Journal of Research in Science Teaching,* 38(6), 631-645.

Lee, O., Buxton, C., Lewis, S., & LeRoy, K. (2006). Science inquiry and student diversity: Enhanced abilities and continuing difficulties after an instructional intervention. *Journal of Research in Science Teaching*, 43 (7), 607-636.

Loucks-Horsley, S., Hewson, P., Love, N., & Stiles, K. (1998). *Designing professional development for teachers of science and mathematics*. Thousand Oaks, CA: Corwin Press

Lynch, S., Kuipers, J., Pyke, C., & Szesze, M. (2005). Examining the effects of a highly rated science curriculum unit on diverse students: Results from a planning grant. *Journal of Research in Science Teaching*, 42(8), 912–946.

Martin-Hansen, L. (2002). Defining inquiry. *The Science Teacher,* 69(2), 34-37.

Maykut, P., & Morehouse, R. (1994). *Beginning qualitative research. A philosophic and practical guide*. London: The Falmer Press.

Miles, M.B., & Huberman, A.M. (1994). *Qualitative data analysis. An expanded sourcebook 2nd ed.* London: Sage Publications.

National Commission on Mathematics and Science Teaching for the 21st Century. (2000). *Before it's too late*. Washington, DC: U.S. Department of Education.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Research Council. (1996a). *National science education standards.* Washington, DC: National Academy Press.

National Research Council. (1996b). *The role of scientists in the professional development of science teachers: Committee on biology teacher inservice programs.* Washington, DC: National Academy Press.

National Research Council. (2000). *Inquiry and the national science education standards.* Washington, DC: National Academy Press.

Newman, W, Abell, S., Hubbard, P., McDonald, J., Otaala, J., & Martini, M (2004). Dilemmas of teaching inquiry in elementary science methods. *Journal of Science Teacher Education*, 15(4), 257-279.

Nurnberger-Haag, J., Huziak-Clark, T., Van Hook, S. J., & Ballone Duran, L. (2008). Mind the gap! A model for collaboration between K-12 teachers and graduate student scientists/mathematicians. *Teaching & Learning: The Journal of Natural Inquiry and Reflective Practice,* 23(1), 20-39.

Ohio Department of Education. (2003). *K-12 Academic content Standards, Science.* Ohio Department of Education.

Pine, J., Aschbacher, P., Roth, E., Jones, M., McPhee, C., Martin, C., Phelps, S., Kyle, T., & Foley, B. (2006). Fifth graders' science inquiry abilities: A comparative study of students in hands-on and textbook curricula. *Journal of Research in Science Teaching*, 43(5), 467-484.

Rutherford, F., & Ahlgren, A. (1990). *Science for all Americans*. Oxford University Press: NY.

Schuster, D. A., & Carlsen, W. S. (2009). Scientists' teaching orientations in the context of teacher professional development. *Science Education*, 93(4), 635-655.

Settlage, J. (2007). Demythologizing science teacher education: Conquering the false ideal of open inquiry. *Journal of Science Teacher Education*, 18(4), 461-467.

Shymansky, J., Kyle, W., & Alport, J. (1983). The effects of new science curricula on student performance. *Journal of Research in Science Teaching,* 20(5), 387-404.

Shymansky, J., Hedges, L., & Woodworth, G. (1990). A reassessment of the effects of inquiry-based science curricula of the 60's on student performance. *Journal of Research in Science Teaching,* 27(2), 127-144.

Songer, N. B., Lee, H., & McDonald, S. (2003). Research towards an expanded understanding of inquiry science beyond one idealized standard. *Science Education*, 87(4), 490-516.

Tafoya, E., Sunal, D., & Knecht, P. (1980*).* Assessing inquiry potential: A tool for curriculum decision makers*. School Science and Mathematics*, 80(1), 43-48.

Wallace, C. S., & Kang, N. (2004). An investigation of experienced secondary science teachers' beliefs about inquiry: An examination of competing belief sets. *Journal of Research in Science Teaching,* 41(9): 936-960.

Wee, B., Shepardson, D., Fast, J., & Harbor, J. (2007). Teaching and learning about inquiry: Insights and challenges in professional development. *Journal of Science Teacher Education,* 18(1), 63-89.

Weiss, I. R., Pasley, J.D., Smith, P.S., Banilower, E., & Heck, D.J. (2003). Looking inside the classroom: A study of K-12 mathematics and science teaching in the United States. Chapel Hill (NC): Horizons Research. Retrieved July 21, 2008, from www.horizon-research.com/insidetheclassroom/reports/looking/.

Welch, W., Klopfer, L., Aikenhead, G., & Robinson, J. (1981). The role of inquiry in science education: Analysis and recommendations. *Science Education*, 65(1), 33-50.

Wise, K., & Okey, J. (1983). A meta-analysis of the effects of various science teaching strategies on achievement. *Journal of Research in Science Teaching*, 20(5), 419-435.

Appendix I: Inquiry Survey (Teacher Version)

1. How would you define learning through inquiry?

2. Describe a lesson where inquiry-teaching methods are being used.

3. What skills do students need to have in order to do inquiry?

4. What skills do teachers need to have in order to teach using inquiry?

5. Describe a classroom environment conducive to inquiry

6. How often did you use inquiry in your classroom this past year? (Example: Once a week, twice a week, once a month, once a quarter)

7. What do you see as the advantage of teaching for inquiry during the upcoming academic year in your classroom?

8. What do you see as the disadvantages of teaching for inquiry during the upcoming academic year in your classroom?

9. Are there any people or groups who would approve or disapprove of your teaching for inquiry during the upcoming academic year in your classroom?

10. What things would encourage you or make it easier for you to teach for inquiry during the upcoming academic year in your classroom?

11. What things would discourage you or make it harder for you to teach for inquiry during the upcoming academic year in your classroom?

12. Do you have any other thoughts or concerns about teaching for inquiry during the upcoming academic year in your classroom?

For the graduate student scientists, questions 7-12 asked about “a future college classroom” instead of “the upcoming academic year in your classroom”.

Appendix II: Graduate Student Scientist Interview Protocol

1. Describe your past K-12 science and math experiences, How did these experiences affect your career plans?

2. Describe your collaboration with your teacher partner—when did you go to the school, when did you plan, how did you plan, how did you implement the plans in your classroom?

3. How has your participation in PRISM impacted your opinion of K-12 science or math? (Follow up—Has your attitude towards teaching and learning science or math in K-12 changed? How or why?)

4. Has your PRISM experience made you want to be more involved in K-12 outreach in your future career?

5. Have your teaching or communication skills improved due to your work as a PRISM Fellow? (Follow up – In what ways have your teaching/communication skills improved?)

6. Has your understanding of math and/or science education improved due to your work as a PRISM Fellow? Please explain.

7. How will you transfer the PRISM teaching experiences into the college setting one day?

8. What were the most important things you learned from your teacher partner? How do you think this will transfer into the college teaching environment?

9. Describe an inquiry lesson that you and your teacher partner taught. What did you learn from this experience?

10. What are your career plans? How has PRISM affected these plans, if at all?

11. Do you believe that your teacher partner has gained the tools necessary to do inquiry in his/her classroom?

12. Do you believe that your teacher partner has gained a greater understanding of math/science and a greater confidence in his/her math/science knowledge?

13. Do you believe that your teacher partner will be able to continue using the lessons and techniques that you developed together next year when you are no longer in his/her classroom?

14. Did you perceive a change in the students over the course of the term (e.g., changes in enthusiasm, interest, confidence)?

15. What effect have you had on the school outside your teacher partner's classroom? For example, what effect do you feel you have had on other teachers, an administrator, etc?

16. Did you have any experiences or learn/synthesize material in the K12 classroom or in preparing exercises for K12 students that benefited your research program or academic studies? If so, please give examples. Did your placement require you to teach topics that were not in your discipline? How did that affect you and your own understanding of science/mathematics?

Trends in Advanced Placement Science and Mathematics Test-Taking Among Female Students in California: A Latent Variable Approach

D. Michael Campbell, Ed.D.
Rossier School of Education
University of Southern California

Richard Brown, Ph.D.
Rossier School of Education
University of Southern California

Elizabeth Perry, M.A.
University of San Diego

Abstract

This study compares trends in participation and performance on all science and mathematics Advanced Placement exams for female and male students in California high schools over a six-year period. Results indicate that while more females are participating in Advanced Placement science and mathematics they are not performing to the levels of their male counterparts. This performance gap presents a real obstacle for females as they prepare to enter college and later compete for jobs in these fields after graduation. As such, these findings signal the need for additional research that identifies means of reducing the performance gap between males and females in Advanced Placement examinations.

Correspondence should be addressed to D. Michael Campbell, Ed.D., 2041 Paseo Dorado #8, La Jolla, Ca 92037, 619-368-5736, cmik731@hotmail.com

Introduction

More than 25 years ago the debate was waged over whether female students were as capable as male students in the subjects of mathematics and science. In the early 1980s, Benbow and Stanley (1980) suggested that there could be a genetic basis for male superiority in mathematics and that females would be better off accepting this differentiation (as cited in Kolata, 1980). Since this time, numerous studies have aimed to address this issue with the purpose of debunking or supporting the work by Benbow and Stanley. More recently, in a 2005 economic conference, Harvard University President Lawrence Summers offered his personal insights on the issue. He remarked that one determining reason for decreased female performance in mathematics and science had to do with “innate ability” in these academic fields. From Summers’ response, a national debate once again erupted over whether intrinsic differences between the sexes were responsible for the underrepresentation of women in mathematics and science.

The renewed debate heighted contemporary scholars’ attention to female participation in mathematics and science. One notable national event occurred when the

American Psychological Association held a 2007 forum entitled *Women in Science: Are They Being Held Back*? Contemporary research has added further attention to this topic with the outlook that motivation (Dumais, 2009; Moody & Linn, 1986; Turner & Harriet, 2003), locus of control (Reis & Park, 2001), testing items (Beller & Gafni, 2000; Walsh, Hickey & Duffy,1999), classroom behaviors (Born, Revelle & Pinto, 2002; Inzlincht & Ben-Zeev, 2002), departmental and institutional factors (Cohoon, 2001), teacher-student interactions (Duffy, Warren & Walsh, 2001; Potvin, Hazari, Tai & Sadler, 2009), teacher preparation (Fraser-Abder, 2001), and self-efficacy (Britner, 2008; Ferla, Valcke & Cai, 2009; Gainor, 1998; Halpern, Benbow, Geary, Gur, Hyde & Gernsbache, 2007b; Herbert & Stipek, 2005; Huebner, 2009; Nauta, Epperson & Kahn, 1998; Pearl, Pollack, Riskin, Thomas, Teshome, Maushak, & Athreya, 2001; Wolf & Wu, 1990) may be useful determinants in leveling the "cognitive field" in these subject areas. While some studies have focused on the biological-based differences and others on learning and socialization issues, neither the biological nor the environmental rationale has produced unequivocal evidence to support the involvement disparities by females (Oakes, 1990). In essence, the competing sides of the literature have yet been able to undeniably prove their positions on why female achievement tends to trail male achievement in mathematics and science.

Decades of research in cognition have largely drawn attention to the fact that female students have seldom participated in mathematics and science in the same numbers as their male counterparts. Most of the contemporary literature is no longer focused on whether female students are as intelligent as their male counterparts; instead, the focus involves examining the motives behind why large numbers of young girls avoid intermediate and advanced mathematics and science curriculum and subsequent careers in associated fields. Numerous difficulties have been illustrated in the literature including teacher intimidation and lack of proper advising (Brainard, Laurich-McIntyre & Carline, 1995; Cooney & Bottoms, 2003), chilly academic climates with messages that reinforce sexists expectations (Ginorio, 1995; Guiso, Monte, Sapineza, & Zingales, 2008), poor mentoring and student relationships with faculty (Herzig, 2004), less student interest and confidence (Catsambis, 1994), dearth of adequate feedback (Halpern, Aronson, Reimer, Simpkins, Star & Wentzel, 2007a; Huebner, 2009), and the deficit of (Nobles & McDonald, 1996) and importance of female role models (Halpern, Aronson, Reimer, Simpkins, Star, & Wentzel, 2007; Huebner, 2009; Karnes & Stevens, 2002). Teacher stereotypes have also been linked to poor female mathematics and science participation. These stereotypes include teacher and textbook sole reliance on prominent male figures in math and science. As Nobles and McDonald (1996) remark, rarely, if ever, were female mathematicians and scientists highlighted as major contributors to these fields. As a consequence, the disparity in mathematics and science participation and upper-level course taking contributes to the large "gap" in the number of females choosing professions in math, science, and technology fields (Campbell, 1992; Johnson, 2000). Furthermore, Dick and Rallis (1991) remarked that even when high school females are performing at higher academic levels than their male counterparts, they continue to express less interest in participating in mathematics and science careers. One theory proposed by Taylor, Friot & Swetnam, (1997) suggests that female choices to pursue mathematics and science education are reinforced daily by individual experiences in and out of the classroom. Many times, female students feel that there is a societal expectation that mathematics and science are "male domains." Given this, Campbell, Jolly, Hoey &

Perlman, (2002) observed that female students are much less apt than male students to continue in career fields in quantitative disciplines. Similarly, Kerr and Robinson-Kurpius (2004) highlight that many Hispanic female students are expected to stay close to home, support family objectives, and adhere to cultural ideals. In the same way, African American young women often lack the social support and educational self-efficacy necessary to persist in math and science majors (Oakes, 1990).

While the literature offers a wide range of reasons why female participation is below that of their male counterparts the notion of stereotype threat has taken new importance in many academic circles. Stereotype threat is a feeling that individuals experience when they are in jeopardy of confirming a negative stereotype in the eyes of others (Spencer, 1997). Research has shown that female students are exceedingly aware that gender stereotypes depict them as being bad at math and science (Bell & Spencer, 2002; Huguet & Régner, 2009; Keller, 2002; Kiefer & Sekaquaptewa, 2007; Marx, Brown & Steele, 1999; Nosek, Smyth, Sriram, Lindner, Devos, Ayala, Bar-Anan, Bergh, Cai, Gonsalkorale, Kesebir, Maliszewski, Neto, Olli, Park, Schnabel, Shiomura, Tulbure, Wiers, Somogyi, Akrami, Ekehammar, Vianello, Banaji, & Greenwald, 2009; Shih, Pittinsky, & Ambady, 2002; Smith, Sansone, & White, 2007; Smith & White, 2002; Steele, 1999; Thoman, White, Yamawaki, & Koishi, 2008; Verity et. al, 2002). As a result, many of these students reason that poor outcomes on mathematics or science tests are directly linked to their gender (Steele, James, & Barnett, 2002). This thinking, in turn, may create anxiety and/or strong performance attributions that may lead to the originally imagined outcomes. To avoid these negative stereotypes, female students may leave mathematics and science courses for more traditionally female options in other fields such as education and the social sciences (Jacobs, 2005).

Fortunately, the contemporary literature has shown that when teachers, especially in middle and high school grades, make math and science classrooms free from stereotype threat (Brownlow, Jacobi & Rogers, 2004), encourage a safe and nurturing environment (Allen, 1995; Gavin, 2000), and remove obstacles that hinder student self-efficacy (Betz & Hackett, 1983; Huang & Brainard, 2001; Kerr & Robinson-Kurpius, 2004; Ziegler & Heller, 2000), female students become more effectively prepared to enter and participate in advanced courses in these fields. Furthermore, parental and mentoring influences have also been shown to positively influence female student preparation and ensuing participation in mathematics and science. Gavin (2000) found that nurturing and encouraging math and science ability through at-home problem solving activities, gender equal career expectations, and exposure to female role models in math and science was a strong foundation of later scholastic ability and reduction of stereotype threat. Moreover, the literature is rich with studies linking nurturing parental (Beckwith, 1983; Ferry, Fouad & Smith, 2000; Gavin, 2000) and mentoring activities (Herzig, 2004; Karnes & Stevens, 2002; Kerr & Robinson-Kurpius, 2004; McLaughlin, 2005; Murphy & Sullivan, 1997) to increased female preparation and stereotype threat reduction in mathematics and science. Whatever the explanation, it is important to identify if the patterns observed in female participation and performance in advanced science and mathematics are persisting, or to what extent they are changing over time.

Objective: Measurement and Assessment of Student Participation and Performance

This study is designed to address two main objectives. The first objective is to compare the extent to which opportunities to take mathematics and science Advanced Placement exams are increasing or decreasing for female students by examining six years of student testing data and to identify features of high schools that relate to greater expansion in Advanced Placement test taking for females in these areas. The second objective is to compare changes in performance on Advanced Placement tests in mathematics and science between male and females students and to identify what features of schools influence these changes in student performance.

Methodology

Whenever we are describing change, the form of change must be identified. Change may be linear-going up or down in a straight line - or it may be nonlinear- going up rapidly then leveling off or accelerating its pace of improvement (Acock & Fuzhong, 1999). For our analyses, we begin with a linear growth model with covariates to explain the rate of change in Advanced Placement test taking and performance. However, in some instances it may be necessary to amend the linear model to include a non-linear component to improve the explanatory capability of the model and improve the model fit to the observed data. In only one case (female mathematics participation) was it necessary to get the model fit below the specified threshold.

The starting point for change over a given time period is referred to as the *intercept (i)***.** The *intercept (i)* is the beginning value of our data set in year 1; identified in Figure I as indicator t_1. Indicators t_1 through t_6 depict the data set years ranging from 1998 to 2003. In addition to the *intercept (i)*, the *linear slope* describes the amount of change for each measured variable. In short, this parameter illustrates how much the curve grows each year (Acock & Fuzhong, 1999). Other variables, called covariates, which are depicted by the indicators $\mathbf{C}_1$ through $\mathbf{C}_4$, may impact the rate of change and consequently give insights into what conditions at the school might relate to varying levels of change.

The school level covariates used in this study include such features as teacher quality, the adult/student ratio, a school's academic momentum, and student achievement level. The covariate teacher quality ***(C_1)*** is measured by the percentage of fully credentialed teachers in each school. The covariate adult/student ratio ***(C_2)*** is measured by the ratio of teachers and certificated staff to students. The covariate academic momentum ***(C_3)*** is a measure of a school's improvement in proficiency rates on state assessment tests. The covariate student test scores ***(C_4)*** is measured by how well students are doing, as measured by their state test scores.

In addition, the relationship between the *intercept (i)*, and the *slope* is estimated in this latent variable growth model, represented by a line starting from the *intercept* and continuing to the slope. Positive values for this relationship reflect faster growth rates for schools with more Advanced Placement testing in the initial year. Negative values reflect slower growth rates for schools with higher levels of student test taking in 1998.

Listed below is a general representation of the Latent Variable Growth Model for males and females in mathematics and science.

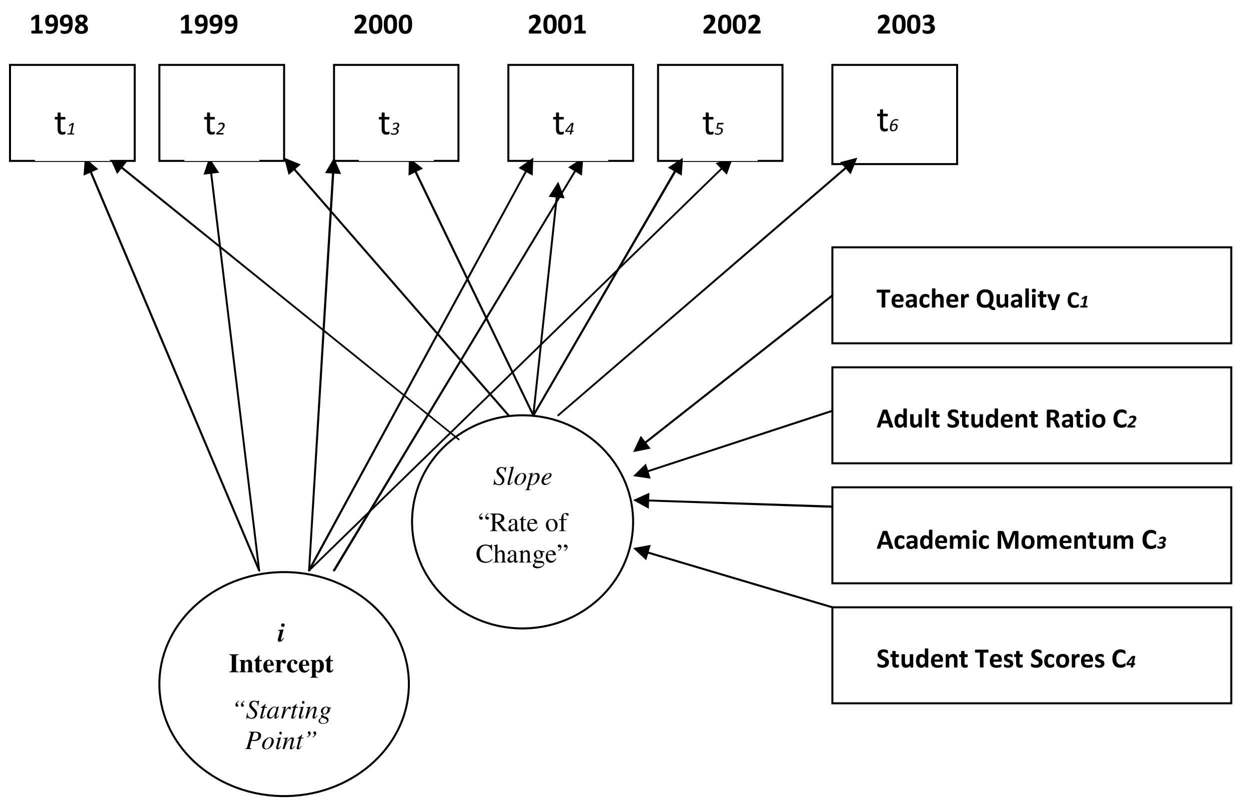


Figure I. General Latent Variable Growth Model

Data Set

The Advanced Placement testing data used in this study include information for all Advanced Placement tests taken by California high school students from 1998 to 2003. In the academic field of science the tests included Biology, Chemistry, Macroeconomics, Microeconomics, Environmental Science, Physics B, and Physics C. In the field of mathematics the tests included Calculus AB, Calculus BC, Statistics Computer Science A, and Computer Science AB. The data were disaggregated by subject area, ethnicity, and school site for all 6 years. Next, the data were disaggregated by gender. The starting data set included 16,383 records from 874 high schools in 12 Advanced Placement subjects.

Results

Model Fit

Before interpreting the estimates derived from the proposed models, it is important to identify how well the models adequately capture the variability in the data. This is usually done through the investigation of a variety of statistical measures of model fit. One measure of model fit is the root mean square error of approximation (RMSEA). This metric ranges from a value of 0 to 1, with lower values indicating better model fit. Several researchers have suggested threshold values for the RMSEA indicating sufficient fit of the model for the data. Hu and Bentler (1999) propose .06 as an optimal critical value for indicating close fit. Browne and Cudeck (1993) suggest values ranging from .06 to .08 indicate acceptable fit, and values ranging from .08 to .10 indicate mediocre fit. RSMEA values above .10 would reflect a poor fitting model. While the linear models using the performance data all indicated very good fit, one linear model (female math participation) using the participation data as outcomes generated a fit statistic beyond the acceptable thresholds. In order to address this and provide a better model fit to the data, a quadratic growth term was added to the model. This allows the growth trajectories to be non-linear, or level off or accelerate over time. This addition resulted in better explanatory power for the model and fit statistic within the acceptable limits.

Participation

The first research question investigated is principally interested in the rate of male and female participation. In order to measure participation, we must focus on the slope segment of this research model. Each slope value illustrates whether the opportunities in mathematics and science Advanced Placement testing are increasing or decreasing for the males and female students examined in this study. The higher the slope value for each group, the faster this group's participation in Advanced Placement mathematics and science is growing. The lower the slope values, the slower each group is increasing participation in Advanced Placement mathematics and science testing. When the slope values are compared in relation to each group, meaningful testing trends can be determined. The growth data that addresses this research question are located in the third column labeled "Slope" in Table I.

Table I
Gender Participation in Mathematics and Science

Subgroup	RMSEA	Intercept	Slope	Teacher Quality	Adult/Student Ratio	Academic Momentum	Student Test Scores
Female Science Participation	0.065	14.40	2.23	NS	0.235	NS	0.379
Male Science Participation	0.065	12.15	2.55	NS	0.196	0.022	0.301
Female Math Participation*	0.084	17.98	3.32	NS	0.252	-.123	0.866
Male Math Participation	0.061	14.48	3.15	NS	0.208	NS	0.500

*included a non-linear term to enhance model fit. The slope reflects the linear slope and the average effect of the non-linear term.

In the discipline of mathematics, males had an intercept of 14.48 and a slope of 3.13. This means that across the full population of high schools in California, the average number of Advanced Placement tests taken by males in mathematics in 1998 was slightly over 14. This group's test participation in math grew at a rate of 3.13 new tests each year or, on average, 3.13 new tests were taken by males each year at each high school. As compared to males, females had a larger intercept value of 17.98 and a slope of 3.32. While the initial value of females was noticeably larger than males, the growth rates are largely comparable. The data showed that gender played a role in the area of mathematics participation as females display higher intercept values (17.98) as compared to their male counterparts (14.48). Additionally, females (3.32) slightly outpaced males (3.13) in mathematics participation growth rates. If these trends continue the data indicate that on average, and across all California high schools, the number of tests taken by females will continue to outpace the number of tests taken by males in mathematics.

In the academic discipline of science, males had an intercept of 12.14 and a slope of 2.55 in participation. This means that across the full population of high schools in California, the average number of Advanced Placement tests taken by males in science in 1998 was slightly over 12. In comparison, females had an intercept of 14.40 and a slope of 2.23. Gender discrepancy was once again noticeable as females, in the area of science participation, displayed higher intercept values (14.40) when compared to their male counterparts (12.15). However, in contrast to mathematics participation, males slightly outpaced female growth rates in science participation. The slope values for both genders (2.55 and 2.23 respectively) in science, as they were in mathematics, are to a large extent similar. See Figure II for a graphical display of the trends in math and science AP testing participation by gender.

The next step in analyzing the data in mathematics and science participation involves investigating the individual covariate effects on the data set. From the examination of the covariate data, the features of high schools that relate to greater expansion in Advanced Placement testing in mathematics and science vary between the genders. In terms of mathematics participation, teacher quality and academic momentum did not have much significance, while adult/student ratio and student test scores were significant for both males and females. This means that the percentage of fully credentialed teachers and a school's improvement in proficiency rates on state assessment tests did not lead to increases in Advanced Placement test taking by males or females. In science participation once again this study found that teacher quality had no significant impact on either group; however, adult/student ratio was shown to relate positively to an expansion in Advanced Placement test taking in science for both males and females. That is, the ratio of adults in the school to students related to positive changes in Advanced Placement science test taking for both males and females. Additionally, academic momentum lead to increased Advanced Placement test taking by males for science only, but had no significant impact on females in science and a negative impact in mathematics.

Performance

In addition to participation, this study also explored the extent to which performance on Advanced Placement tests in mathematics and science improved or lessened for each gender and sought to identify whether specific features of schools influenced student performance. Rather than the number of students in each group taking a given Advanced Placement test, the outcome measure modeled for these analyses is the proportion of students passing the Advanced Placement tests of interest with a grade of "3" or better. Although the models are similar, the outcome measures are distinctly different from the earlier analysis. Parameter estimates and measures of model fit for each model are summarized in Table II.

Table II
Gender Performance in Mathematics and Science

Subgroup	RMSEA	Intercept	Slope	Teacher Quality	Adult/Student Ratio	Academic Momentum	Student Test Scores
Female Science Performance	0.039	0.316	0.007	NS	NS	NS	0.008
Male Science Performance	0.035	0.401	0.008	NS	NS	NS	0.009
Female Math Performance	0.048	0.410	0.009	NS	NS	NS	0.008
Male Math Performance	0.051	0.486	0.011	NS	NS	NS	0.005

In the academic discipline of mathematics, males and females had intercepts values of .486 and .410 respectively. This means that the percentage of male and female students scoring “3” or better, on average, across all California high schools, was 48.6% for males and 41.0% for females. The slopes for males and females were relatively similar with values at .011 and .009 respectively. A data value of .011 indicates that the percentage of male students passing Advanced Placement tests increased by a value of 1.1 each year. Likewise the percentage of female students scoring “3” or better grew by a value of .9 each year. The data suggest that gender groups differed in the area of mathematics performance as males (.486, .011) displayed higher intercept and slope values as compared to their female (.41, .009) counterparts.

In the academic discipline of science, males (.401) had higher intercept values in science performance as compared to females (.316). This means that across the full population of high schools in California, on average, the percentage of male and female students scoring “3” or better on Advanced Placement tests was 40.1% and 31.6% respectively. The male student slope value of .008 indicates that for every year, the average increase in the percentage of male students scoring “3” or better was .8%. In contrast to males, females had a slope of .007; meaning that, on average, an additional .7% of female students scored “3” or better in Advanced Placement test taking per year. While these data are not that widely divergent, it is worth noting that in both mathematics and science performance, males began with higher passing relates and continue to outpace females in performance rates throughout the data years 1998 to 2003. The data indicate that the performance gap between males and females is not reducing. See Figure III for a graphical display of the trends in math and science AP testing performance by gender.

In terms of the covariate data, teacher quality was not shown to positively relate to the increases in the percentage of males or females passing Advanced Placement examinations in mathematics and science. This finding suggest that teacher quality in California high schools had no positive affect on the percentage of males or females passing Advanced Placement examinations in these disciplines. Similarly, the variables adult/student ratio and academic momentum failed to show a positive relationship to increased percentages of males and females scoring “3” or better on Advanced Placement examinations in math and science. Only academic achievement at the schools, as measured by performance on state examinations, was positively related to an increased percentage in both genders passing Advanced Placement tests in mathematics and science. This suggests that greater increases in passing rates for males and females occur in high schools with better academic achievement.

Study Implications

The prominent gender distinctions are centered on the fact that females took more tests in both mathematics and science compared to their male counterparts. These findings are significant because they show that females are, in fact, actively involved in mathematics and science Advanced Placement test taking. The data suggest that females are participating at comparable rates in mathematics and science Advanced Placement testing and are continuing to make substantial strides in this area. The data also indicates that the growth rates for each gender were affected by certain features in California high schools that relate to greater expansion in Advanced Placement test taking in mathematics and science. These features include adult/student ratio, a school's academic momentum, and student test scores on state examinations.

In the field of mathematics, the findings show that the adult/student ratio and student test scores largely influenced the growth rates for both male and female students. However, neither male nor female growth rates were influenced by teacher quality or a high school's academic momentum. The data findings are important because they show that greater expansion in Advanced Placement testing will occur for males and females in high schools with positive adult/student ratios and strong academic achievement.

In mathematics and science participation, the data demonstrated that females had higher initial participation rates as compared to their male counterparts. Females also grew at higher or comparable rates in mathematics and science over the years. These findings show that in terms of Advanced Placement test taking, females are matching or outpacing males in the areas of mathematics and science. The data suggest that females, in fact, are actively participating in Advanced Placement test taking in mathematics and science.

Males had higher intercept and slope values in both mathematics and science performance data. This means that males, as compared to females, have higher initial passing scores in Advanced Placement test taking in mathematics and science. Furthermore, males demonstrate marginally stronger yearly performance growth rates in mathematics and science Advanced Placement test taking. When these data are contrasted with the data in gender participation, it illustrates that although more females are participating in Advanced Placement mathematics and science test taking their passing rates are still lower than that of males. Although female participation rates in mathematics and science are encouraging, the implications for this group's performance data are problematic. While more females are participating in Advanced Placement mathematics and science, they are not performing at the levels of their male counterparts. This performance gap presents a real obstacle for females as they prepare to enter college and later compete for jobs in these fields after graduation. Furthermore, the gap is not diminishing. Ergo, the problem self-perpetuates as reduced numbers of females in mathematics and science lead to fewer successful examples of mentors for aspiring female students. Ultimately, the findings suggest that continued underrepresentation in the number of females with mathematics and science majors in colleges and universities will lead to decreased participation of females in the domestic workforce. This reduction in the workforce may lead to a continued lack of economic and social power for this group.

Conclusions

Understanding participation and performance in Advanced Placement mathematics and science is exceedingly important for California's future. As greater numbers of females are making up the populations of California's elementary, middle, and high schools, adequate and equitable mathematics and science preparation for these students is essential. This research identifies patterns of participation and performance of female high school students in California on Advanced Placement mathematics and science examinations over a five-year period of time. Rates of participation in Advanced Placement testing indicate that female students are participating nearly on par with their male counterparts and even at greater rates in mathematics; however, the identified lower performance of females on Advanced Placement examinations is a cause for concern.

While this study reveals that academic achievement, as measured by test scores on state examinations, correlates with greater performance on Advanced Placement examinations, more research is needed to examine why female students are not performing as well as male students. As gender equity in participation increases, what explains the disparity in performance levels? Are instructional strategies unequal or insufficient, leading to lower performance by females? Or, for example, are Advanced Placement tests an inaccurate or inequitable assessment of skill, due to implicit stereotyping or other gender bias? There is likely to be more than one factor at play. Identifying why female students in California underperform on Advanced Placement testing in mathematics and science in relation to male students is imperative for increasing female involvement in these fields in higher education and related professions. What message do low-test scores send to females about ability – and in turn, how does this influence the decisions of female students to pursue academic degrees and professions in the fields of mathematics and science.

Once the causes of underperformance are determined, further research is needed to identify the means for reduce this disparity. A recent study by Halpern et al. (2007) identifies five strategies to encourage females in mathematics in science, including: teaching females students that success in mathematics and science is not based on innate ability, increasing exposure of female students to successful female mathematicians and scientists, providing "prescriptive, informational feedback," creating classroom environments that engage and create lasting interest in science and math, and providing additional training for female students in spatial skills (p. 6). Identifying why female students are underperforming in relation to their male classmates will aid in identifying if Halpern et al.'s suggested strategies suffice, or if additional or modified strategies are needed to address test performance specifically.

There are a multitude of reasons why Advanced Placement test taking should be equitable for all students. The two important reasons include the competitive nature of college admissions for in-state colleges and universities and the fact that students are able to decrease the cost of college attendance by earning college credit for these courses. Another important reason is that there have been decades of inadequate preparation that have created a widening deficit of qualified workers in the global workplace. In essence, as fewer numbers of qualified females focusing on mathematics and science enter the educational pipeline, fewer students from this group believe careers in mathematics and sciences are obtainable. This process leads to a lack of creative ingenuity and decreased domestic competitiveness in the global industrial workplace.

A substantial portion of the current literature on mathematics and science preparation (Brainard, Laurich-McIntyre, & Carline, 1995; Brown, 2004; Furry and

Beasley, 1999; Furry and Hecsh, 2001; Johnson, 2000; Oakes, 1990; Solorzano & Ornelas, 2002; Stanley, 1997; Rinne, 2000, and Ratliff, 2001) discusses the declining trends of female participation in mathematics and science and how these trends could lead to insufficient representation in industry and educational leadership positions. However, few studies have looked beyond female participation in advanced mathematics and science testing to examine the performance rates of these students. While it is important to inspect the number of tests taken by gender, it is also essential to examine the percentage of students passing Advanced Placement tests in mathematics and science and their rates of change over time. The significance of this study rests in the fact that it highlights the growth of both genders, what school features impact growth rates, and then observes exactly where the discrepancies are occurring.

References

Acock, A. C., & Fuzhong, L. (1999, October). *Latent Growth Curve Analysis: A Gentle Introduction.* Paper presented at the Oregon State University Department of Human Development and Family Sciences conference, Corvallis, OR.

Allen, D. (1995). Encouraging success in female students: Helping girls develop math and science skills. *Gifted Child Today, 18*(20), 4-45.

Beckwith, J. (1983). Gender and math performance: Does biology have implications for educational policy? *Journal of Education, 165*(2), 158-174.

Benbow, C. P. & Stanley, J. C. (1980). Consequences in high school and college of sex differences in mathematical reasoning ability: A longitudinal perspective. *American Education Research Journal, 19*(4), 598-622.

Bell, A.E., & Spencer, S. J. (2002). The effect of stereotype threat on women's performance on the fundamentals of engineering exam. Proceedings of the 2002 American Society for Engineering Education Annual Conference & Exposition. Retrieved December 18, 2007 from http://www.ecpe.vt.edu/fac_support/DSPCL/docs/ASEE02.pdf

Beller, M. & Gafni, N. (2000). Can item format account for gender differences in mathematics achievement? *Sex Roles, 42*(2), 1-21.

Betz, N. E. & Hackett, G. (1983). The relationship of mathematics self-efficacy expectations to the selection of science-based college majors. *Journal of Vocational Behavior, 23*(3), 329-345.

Born, W. K., Revelle, W., & Pinto, L. H. (2002). Improving biology performance with workshop groups. *Journal of Science Education and Technology, 11*(4), 347-365.

Brainard, S. G., Laurich-McIntyre, S., & Carline, L., (1995). Retaining female undergraduate students in engineering and science. J*ournal of Women and Minorities in Science and Engineering, 2*(4), 255-267.

Britner, S. L. (2008). Motivation in high school science teachers: A comparison of gender differences in life, physical, and earth science classes. *Journal of Research in Science Teaching*, *45*(8), 955-970.

Brown, R. (2004). Changes in Advanced Placement Test taking in California 1998-2003. (Draft deliverable to UC All Campus Consortium On Research for Diversity). Los Angeles, CA: University of California, UC ACCORD).

Browne, M.W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Brownlow, S., Jacobi, T., & Rogers, M. I. (2004). Science anxiety as a function of gender and experience. *Sex Roles*, *42(1-2)*, 1573-2762.

Campbell, P. B. (1992). *Nothing can stop us now: Designing effective programs for girls in math, science, and engineering*. New York, NY: Women's Educational Equity Act Publishing Center.

Campbell, P. B., Jolly, E., Hoey, L., & Perlman, L. K. (2002). *Upping the numbers: Using research-based decision making to increase diversity in the quantitative disciplines* (College Rep. No. 02-02). Princeton, NJ: GE Foundation.

Catsambis, S. (1994). The path to math: Gender and racial-ethnic differences in mathematic participation from middle school to high school. *Sociology of Education*, 67(1), 199-215.

Cohoon, J. M. (2001). What causes women to discontinue pursuing the undergraduate computer science major at higher rates than men: Toward improving female retention in the computer science major. *Communications of the ACM 44(*5) 108-114.

Conney, S., & Bottoms, G. (2003). *Middle grades to high school: Mending a weak link* (Report No. EA-032-691). Atlanta, GA: Southern Regional Education Board. (ERIC Reproduction Service N. ED479785)

Dick, T. P., & Rallis, S. F. (1991). Factors and influences on high school student's career choices. *Journal of Research in Mathematics Education, 22*(4), 281-292.

Duffy, J., Warren, K., & Walsh, M. (2001). Classroom interactions: Gender of teacher, gender of student, and classroom subject. *Sex Roles, 45*(9/10), 579-593.

Dumais, S. A. (2009). The academic attitudes of American teenagers, 1990-2002: Cohort and gender effects on math achievement. *Social Science Research*, 38(4), 767-780.

Ferry, T. R., Fouad, N. A., & Smith, P. L. (2000). The role of family context in a social cognitive model for career-related choice behavior: A math and science perspective. *Journal of Vocational Behavior, 57*(1), 348-364.

Ferla, J., Valcke, M., & Cai, Y. (2009). Academic self-efficacy and academic self-concept: Reconsidering structural relationships. *Learning and Individual Differences*, *19*(4), 499-505.

Fraser-Abder, P. (2001). Preparing science teachers for culturally diverse classrooms. *Journal of Teacher Education, 12*(2), 123-131.

Furry, W.S., & Beasley, K. (Eds.) (1999). The Advanced Placement program: California's 1997-98 experience. Sacramento, California: California State University, Sacramento, The CSU Institute for Educational Reform.

Furry, W.S., & Hecsh, J. (Eds.) (2001). *Characteristics and performance of Advanced Placement classes in California*. Sacramento, California: California State University, Sacramento, The CSU Institute for Educational Reform.

Gainor, K. A. (1998). Social cognitive expectations and racial identity attitudes in predicting the math choice intentions of Black college students. *Journal of Counseling Psychology, 45*(4), 403-413.

Gavin, K. M. (2000). *What parents need to know about…encouraging talented girls in mathematics*. Paper presented at the annual meeting for the National Research Center on the Gifted and Talented, New Haven, CT.

Ginorio, A. B. (1995). *Warming the climate for women in academic science*. Washington, DC: Association of American Colleges and Universities.

Guiso, L., Monte, F., Sapienza, P., Zingales, L., (2008). Culture, gender, and math. *Science, 320*(5880), 1164-1165.

Halpern, D. F., Aronson, J., Reimer, N., Simpkins, S., Star, J., & Wentzel, K. (2007a). *Encouraging girls in math and science: IES practice guide* (NCER 2007-2003). Washington, DC: Institute of Educational Sciences, U.S. Department of Education. Retrieved November 4, 2009 from http://ies.ed.gov/ncee/wwc/pdf/practiceguides/20072003.pdf

Halpern, D. F., Benbow C. P., Geary, D. C., Gur, R. C., Hyde, J.S., & Gernsbache, M. A. (2007b). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest*, 8, 1-51.

Haung P. M., & Brainard, S. G. (2001). Identifying determinants of academic self-confidence among science, math, engineering, and technology students. *Journal of Women and Minorities in Science and Engineering, 7*(4), 315-337.

Herbert, J. & Stipek, D. (2005). The emergence of gender differences in children's perceptions of their academic competence. *Applied Developmental Psychology*, 26(3), 276-295.

Herzig, A. H. (2004). Slaughtering this beautiful math: graduate women choosing and leaving mathematics. *Gender and Education, 16*(3), 379-395.

Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.

Huebner, T. A. (2009). Encouraging girls to pursue math and science. *Educational Leadership*, *67*(1), 90-91.

Huguet, P. & Régner, I. (2009). Counter-stereotypic beliefs in math do not protect school girls from stereotype threat. *Journal of Experimental Social Psychology*, *45*(4), 1024-1027.

Inzlincht, M. & Ben-Zeev, T. (2002). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science, 11*(5), 365-371.

Jacobs, J. E. (2005). *Twenty-five years of research on gender and ethnic differences in math and science career choices: What have we learned*? New York, NY: Wiley Periodicals.

Johnson, R. M. (2000, March). *Gender differences in mathematics performance: Walberg's educational productivity model and the NELS: 88 Database*. Paper presented at the meeting of the American Educational Research Association, New Orleans, LO.

Karnes, F. A., & Stephens, K. R. (2002). *Young women of achievement: A resource for girls in science, math and technology.* Amherst, NJ: Prometheus Books.

Keller, J. (2002). Blantant stereotype threat and women's math performance: Self-handicapping as a strategic means to cope with obtrusive negative performance expectations. *Sex Roles, 47*(3/4), 1893-198.

Kerr, B., & Robinson-Kurpius, S. E. (2004). Encouraging talented girls in math and science: effects of a guidance intervention. *High Ability Studies, 15*(1), 85-102.

Kiefer, A.K., Sekaquaptewa, D. (2007). Implicit stereotypes, gender identification, and math-related outcomes: A prospective study of female college students. *Psychological Science*, *18*(1), 13-18.

Kolata, G.B. (1980). Math and sex: Are girls born with less ability? *Science, 210*(4), 1234-1235.

Marx, D. M., Brown, J. L., & Steele, C. M. (1999). Allport's legacy and the situational press of stereotypes. *Journal of Social Issues, 55*(3), 491-502.

McLaughlin, R. (2005). Girls in science. *Science Scope, 28*(7), 14-15.

Moody, C. D. & Linn, E. (1986). *Fostering sex equity in math*. Ann Arbor, MI: McGraw-Hill.

Murphy, D. S., & Sullivan, K. (1997). *Connecting adolescent girls of color and math/science interventions*. (Report No. SE-060-479). Baltimore, MD: Association of Supervision and Curriculum Development. (ERIC Document Reproduction Service No. ED410106)

Nauta, M. M., Epperson, D. L., & Kahn, J. H. (1998). A multiple-groups analysis of predictors of higher level career aspirations among women in mathematics, science, and engineering majors. *Journal of Counseling Psychology, 45*(4), 483-496.

Nobles, C. H., & McDonald, S. B. (1996). Equity under the microscope: Forgotten women in math and science. *School Library Journal, 42*(1), 28-31.

Nosek, B. A., Frederick, L. S., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., Bar-Anan, Y., Bergh, R., Cai, H., Gonsalkorale, K., Kesebir, S., Maliszewski, N., Neto, F., Olli, E., Park, J., Schnabel, K., Shiomura, K., Tulbure, B. T., Wiers, R, W., Somogyi, M., Akrami, N., Ekehammar, B., Vianello, M., Banaji, M. R., Greenwald, A. G. (2009). National differences in gender-science stereotypes predict national six differences in science and math achievement. *Proceedings of the National Academy of Science*, *106*(26), 10593-10597.

Oakes, J. (1990). Multiplying inequalities: The effects of race, social class, and tracking on opportunities to learn mathematics and science (Report No. NSF-R-3928). Washington, D.C.: National Science Foundation. (ERIC Document Reproduction Services No. RIEJUL1 991)

Pearl, A., Pollack, M. E., Riskin, E., Thomas, B., Wolf, E., & Wu, A. (1990). Becoming a computer scientist: A report by the ACM committee on the status of women in computing science. *Communication of the ACM, 33*(11), 47-57.

Potvin, G., Hazari, Z., Tai, R. H. & Sadler, P. M. (2009). Unraveling bias from student evaluations of their high school science teachers. *Science Education*, 93(5), 827-845.

Ratliff, C. (2001). *Performance indicators of California higher education*. (Report No. HE-034-055). Sacramento, CA: California State Postsecondary Education Commission. (ERIC Document Reproduction Service No. ED453731)

Reis, S. M., & Park, S. (2001). Gender differences in high-achieving students in math and science. *Journal for the Education of the Gifted, 25*(1), 52-73.

Rinne, D. H. (2001*). Advanced Placement exams 1999-2000: Results for WCPSS and EBN districts* (Report No. 01.35). Raleigh, NC: Department of Evaluation and Research. (ERIC Document Reproduction Service No. ED457227)

Shih, M., Pittinsky, T. L., & Ambady, N. (2002). Stereotype susceptibility: Shifts in quantitative performance from socio-cultural identification. *Sex Roles, 34*(1/2), 334-335.

Smith, J. L., Sansone, C., & White, P.H. (2007). The stereotypes task engagement process: The role of interest and achievement motivation. *Journal of Educational Psychology*, *99*(1), 99-114.

Smith, J. L., & White, P. H. (2002). An examination of implicitly activated, explicitly activated, and nullified stereotypes on mathematical performance: It's not just a woman's issue. *Sex Roles, 47*(3/4), 179-191.

Solorzano, D.G., & Ornelas, A. (2002). A critical race analysis of Advanced Placement classes: A case of educational inequalities. *Journal of Latinos and Education, 82*(1), 215-229.

Spencer, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52*(6), 613-629.

Stanley, J.C. (1997, April). *Gender differences, especially on Fifty College Board Achievement Tests.* Paper presented at the Annual Meeting of the Eastern Psychological Association, Washington D.C.

Steele, C. M. (1999). Thin ice: Stereotype threat and Black college students. *The Atlantic Online, 22*(2), 22-34.

Steele, J., James, J. B., & Barnett, R. C. (2002). Learning in a man's world: Examining the perceptions of undergraduate women in male-dominated academic areas. *Psychology of Women Quarterly, 26*(1), 46-50.

Taylor, M. J., Friot, F. E., & Swetnam, L. (1997). *Women who say "yes" when science says "no": their lessons for future students and teachers* (Report No. SP-037-223). Phoenix, AZ: American Association of Colleges for Teacher Education. (ERIC Document Reproduction Service No. ED405340)

Teshome, Y., Maushak, N., & Athreya, K. (2001). Attitude toward informal science and math: A survey of boys and girls participating in hands-on science and math. *Journal of Women and Minorities in Science and Engineering, 7*(1), 59-74.

Thoman, D. B., White, P. H., Yamawaki, N., Koishi, H. (2008). Variations of gender-math stereotype content affect women's vulnerability to stereotype threat. *Sex Roles*, *58*(9-10). 702-712.

Turner, R. C., & Harriet, L. A. (2003). Gender differences in cognitive and noncognitive factors related to achievement in organic chemistry. *Chemical Education Research, 80*(5), 563-568.

Verity, P.G., Gilligan, M. R., Frischer, M. E., Booth, M. G., Richardson, J. P., & Franklin, C. (2002). Improving undergraduate research experiences: Lessons from a historically black university's unusual collaboration. AAHE Bulletin, 3-6.

Walsh, M., Hickey, C., & Duffy, J. (1999). Influence of item content and stereotype situation on gender differences in mathematical problem solving. *Sex Roles, 41*(3/4), 219-240.

Ziegler, A., & Heller, K. A. (2000). Conditions for self-confidence among boys and girls achieving highly in chemistry. *Journal of Secondary Gifted Education*, 11(3), 144.

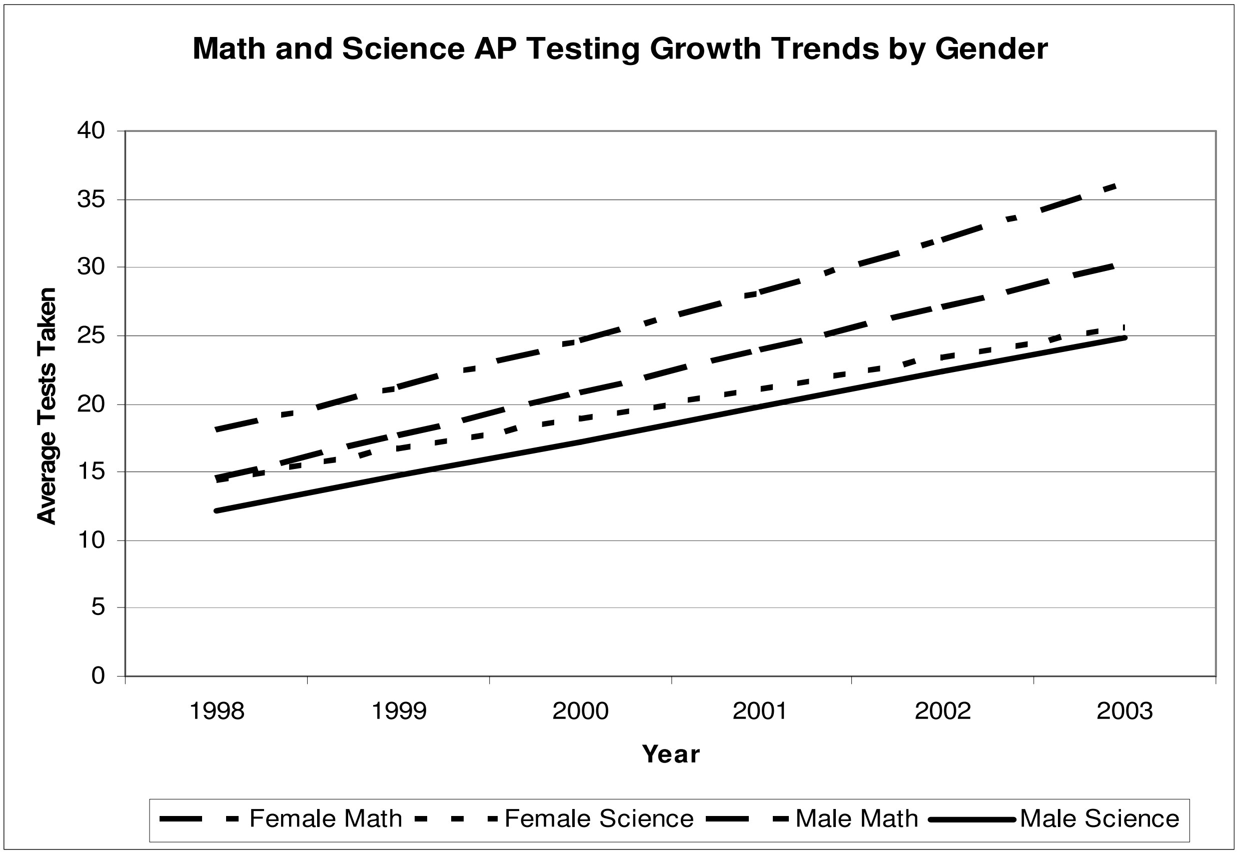
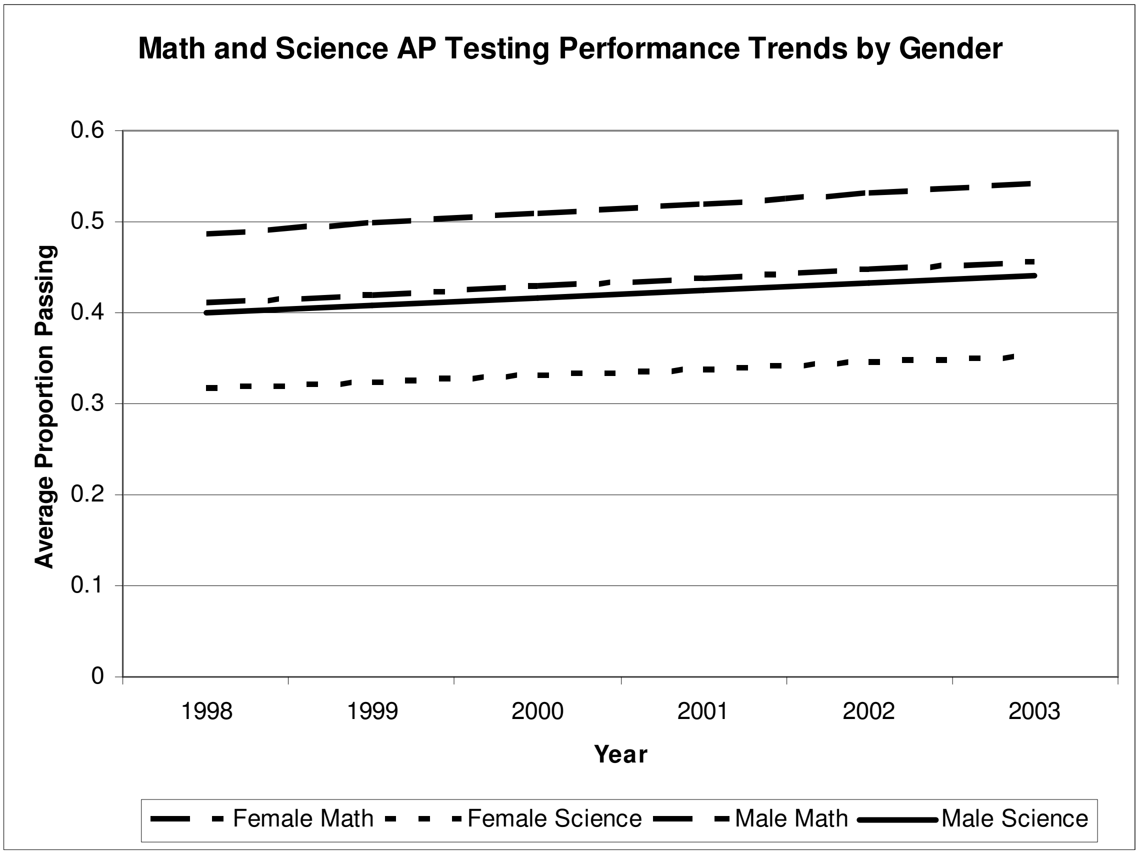

Math and Science AP Testing Growth Trends by Gender
Average Tests Taken
40
35
30
25
20
15
10
5
0
1998
1999
2000
2001
2002
2003
Year
Female Math
Female Science
Male Math
Male Science

Figure II.



Math and Science AP Testing Performance Trends by Gender
Average Proportion Passing
0.6
0.5
0.4
0.3
0.2
0.1
0
1998
1999
2000
2001
2002
2003
Year
Female Math
Female Science
Male Math
Male Science

Figure III.

Student Success in Recognizing Definitions of Eight Terms Found in Fourth Grade Science Textbooks

Robert E. Yager
University of Iowa

Hakan Akcay
Marmara University

Aeran Choi
Kent State University

Stuart O. Yager
Bethel College

Abstract

Continued Professional Development (PD) efforts for science teachers in Iowa have occurred over the 1982 – 2004 years. Teachers have comprised over half of the staff in the PD program while also being partners with action research projects. This is a study of student recognition of key terms across 4th, 8th, and 12th grades for classes taught by a team of the teachers at five year intervals over the 1985 through 2000 academic years. Results indicate that there is no increasing success with such recognition over the grade levels sampled or any major changes over time. But, interestingly more use of the NSES and more focus on real world contexts for science study did not result in any less recognition of the science vocabulary words selected.

Correspondence should be addressed to Robert E. Yager, University of Iowa at Robert-Yager@uiowa.edu

A major research project supported by the National Science Foundation and conducted at the University of Northern Illinois was concerned with factors affecting an "attentive public" regarding science and technology. The research was headed by Jon Miller, a political scientist, and assisted by a sociologist and a science educator (Miller, Suchner & Voelker, 1980). Science attentive persons were defined as persons exhibiting 1) interest in science, 2) knowledge of science, and 3) the ability to increase both. To assess the knowledge dimension the researchers chose four science terms for use with a large national sample of 3000 secondary school students (grades 10-12). The terms selected were molecule, amoeba, DNA, and organic chemical. A striking finding was the fact that secondary students showed no growth across high school grade levels in defining the selected terms (Voelker, 1982). Further, the percent of students who were able to demonstrate mastery was unexpectedly low. For example, only 40% of the students could demonstrate knowing and understanding "amoeba" and "DNA"; 25% could demonstrate knowing the term "molecule"; and only 20% could demonstrate understanding the term "organic chemical".

Past research has focused upon excessive use of words and terminology in science teaching as well as using them as indicators of student learning (Hurd, Robinson, McConnel & Ross, 1981; Yager, 1983; Yager & Yager, 1985). Often science classes have been places where new (and strange) words are taught as the major outcomes of instruction and used as major indicators of student achievement and as evidence of teacher success. Teachers and other leaders often start their planning with the assumption that learners first need the technical vocabulary before they can do science! Interestingly, the focus on new and technical words in science classes surpasses the introduction of new vocabulary in foreign language classes. Often examinations in science stress the mastery of specific terms (Stake & Easley, 1978). The Project Synthesis research reported that 90% of all K-12 science teachers emphasize only mastery of science content (largely by remembering terms) in excess of 90% of instructional time (Harms & Yager, 1981).

K-12 science curricula remain mainly focused on conceptual and factual information. Most teachers and students view science as a body of facts, a collection of formulas, and directed problem-solving methods, all disconnected from the daily lives of students. Evidence of learning is too often based on memorization (NRC, 2007; Seymour & Hewitt, 1994; Trumbull & Kerr, 1993). Most of the typical achievement measures in science during the 80s depended on a special and/or technical vocabulary; there was also great focus in science textbooks upon science terminology and italicized words. At times learning seemed wholly dependent upon mastery of such special science terms. And yet, linguists concerned with vocabulary per se insist that terms are meaningless unless there is first meaning and use established for them (Dale, 1962).

Terminology is best learned when there is a need – often to explain some complex structure or phenomenon. A recent National Research Council report indicates that current approaches for science instruction for young and novice learners may actually be counterproductive (NRC, 1996, p 13). For example, limiting them to learning about discrete science facts without opportunities for discussion, reflection, or direct investigation of the phenomena can lead to a very impoverished understanding of the ideas. Developing expertise in science means developing a rich interconnected set of concepts that move closer and closer to resembling the structure of knowledge in science disciplines found in colleges. Memorizing lists of established scientific facts does not provide the kind of engagement with ideas that will produce rich and interconnected knowledge nor does it help students reason (NRC, 2007, p 338).

Recently Marzano (2004, 2009) has reviewed the research dealing with vocabulary instruction. He maintains that teaching vocabulary improves if a six step process is used. These six are: 1) provide a description, explanation, or example of the new term, 2) ask students to restate the description, explanation, or example in their own words, 3) ask students to construct a picture, pictograph, or symbolic representation of the term, 4) engage students periodically in activities that help them add to their knowledge of the terms in their vocabulary notebooks, 5) periodically ask students to discuss the terms with one another, and 6) involve students periodically in games that enable them to play with terms. Such instructional protocols reach beyond teaching and learning of special vocabulary in isolation.

In many respects the Iowa teachers provided real contexts and classroom practices that were/are the STS features central to Chautauqua, SS&C, and Title IIa projects. These features not unlike the teaching suggested by Marzano include: 1) student identification of problems with local interest and impact, 2) use of local resources (human and material) to locate information that can be used in problem resolution, 3) active involvement of students in seeking information that can be applied to solve real-life problems, 4) extension of learning beyond the class period, the classroom, the school, 5) focus on the impact of science and technology on individual students, 6) viewing science content as more than concepts which exist for students to master on tests, 7) emphasis on process skills which students can use in their own problem resolution, 8) emphasis on career awareness – especially careers related to science and technology, 9) opportunities for students to experience citizenship roles as they attempt to resolve issues they have identified, 10) identification of ways that science and technology are likely to impact the future, and 11) experiencing some autonomy in the learning process (as individual issues are identified). (NSTA, 2008-09, pp. 242-243)

Marzano's work puts a different spin on learning technical terminology. Instead it is how teachers teach and how they involve students in thinking and actions rather than assessing what they are asked to memorize, often by merely recognizing correct definitions. The NSTA view of STS instruction is similar.

Certainly the reform efforts from the 1980s and beyond have focused on other aspects of science, especially on the process skills scientists use to increase understanding of the objects and events in the natural universe. One elementary program, Science – A Process Approach (SAPA), identified 13 skills used by scientists and organized them as the focus for a whole K-8 program (AAAS, 1965). The SAPA program influenced other elementary science curricula and textbook publishers by focusing on general procedures rather than helping students build frameworks of integrated science concepts and processes (NRC, 2007). Students were asked to perform many science activities, making observations and reporting measurements without understanding what they did nor why they did it. Understandingly students often fail to develop meaningful understandings as a result of such programs. Another criticism of SAPA (a process only focus) is that it was based solely on developmental assumptions about student reasoning and learning capacities (Metz, 1995; NRC, 2007).

Generally students enjoyed "learning" the skills – but often their use in any other contexts did not occur. Few attempts to *unite* science concepts with processes were undertaken until the development and release of the National Science Education Standards (NRC, 1996). The NSES reported this "unification" as one of eight facets of science content – and perhaps the most important.

It is clear that the NSES as released in 1996 was much about de-emphasizing special science vocabulary while also providing a rationale for discontinuing the major focus on mastery of basic "discipline-bound" concepts. There also is now a new focus on not considering concepts and/or processes singly and without meaningful contexts. This situation led to the research central to this study. Are there changes in student abilities to know and understand basic terms over the years following the 1983 and 1985 studies (Metz, 1995; Yager, 1983; Yager & Yager, 1985)?

This study was possible because of the longevity of professional development efforts for PreK-12 teachers in Iowa, beginning in 1983 with the Iowa Chautauqua Program (later merging with the $4 million Scope, Sequence, and Coordination (SS&C) project which was one of the six state efforts also coordinated by the National Science Teachers Association (NSTA). The Chautauqua program was funded in Iowa following the NSF funding for it through 1998 followed by three Title IIa projects utilizing the same format and focus on instruction. This means data collected from Iowa teachers enrolled in the programs from the first efforts in 1982 to the current efforts in the international arena. One unique facet of the Iowa professional development program is the involvement of teachers as full partners in heading the summer workshops and the annual follow-up short courses (3 day workshops in October and a second in April) – and often over a three-year interim. All staff (and teachers) enrolled are routinely engaged in multiple action research projects. Also of importance is the fact that both Chautauqua and SS&C were assessed and approved by the U.S. Department of Education's Program Effectiveness Panel (PEP) which was the precursor for funding and dissemination to other states and regions as part of the U.S. Department of Education's National Diffusion Network (NDN).

The Iowa programs encourage teacher involvement in Action Research. One of these collaborative efforts was a follow-up of the Miller, Suchner, & Voelker (1980) work dealing with mastery of textbook terms. It was common to include the same terms used in the early work. Other features of the Iowa professional development effort included the Science-Technology-Society reforms that utilized the six domains for teaching, learning, and assessment (Yager, 1996). This research has focused on concept and process mastery, creativity and attitude as "enabling" domains, a major focus on application of concepts and skills in new situations – inclusion of technology (the human-made world) as well as pure science, and finally, a focus on the history and nature of science. These are all important and tend to de-emphasize the major focus in most K-16 classrooms as curriculum structures that characterize textbooks and state standards. They also focus less on terminology per se.

This larger context is mentioned to give more reason and a setting for a study looking at concept mastery as related to recognition of accurate definitions of eight terms. Those used by Miller, Suchner, & Voelker (1980) and five others used early in Professional Development projects in Iowa provided the sample of terms used in this study. Actually as many as 50 other terms for varying concepts have been used by groups of teachers over varying time frames. They have produced other results indicating the fate of less focus on such terminology – even when it was not a primary focus of teaching in the Iowa professional development programs. State testing and historical focus for most traditional teaching of science was affected by the "vocabulary first" idea. The use of such terminology in introductions for textbook chapters that define the curriculum seem to continue in spite of the many current reforms efforts.

The eight terms were the ones selected by Teacher Leaders. This study provides a look at changes in student performances over time. Although this study is focused on data collected from PEP reports, NDN experiences, and reports to NSF officials, it is used here to illustrate what has happened over the fifteen year period with respect to recognizing accepted definitions of eight science textbook terms.

Specific Research Questions for this study are:

1) How do 4^{th}, 8^{th}, and 12^{th} grade students compare when selecting the most accurate definitions for eight terms found in 4^{th} grade textbooks?

2) How does the ability of students to recognize such “correct” definitions change over a 15 year interim?

3) How do the findings affect and/or negate the reform agenda indicated in the National Science Education Standards? (As illustrated by less and more emphasis conditions related to teaching, PD, assessment, and content)?

Methodology

The research instrument was developed by selecting eight science terms – 3 from the original Miller, Suchner, and Voelker study in 1980 and also included in a 1985 study involving teachers in one large school district in Iowa where no teachers were enrolled in any long range professional development programs (Yager & Yager, 1985). Teacher teams helped develop the four distractors for the multiple choice items which were selected from a typical fourth grade textbook. The distractor items came largely from students who shared their misconceptions in actual classrooms.

Some teachers used the misconceptions to plan additional learning activities. Some of these efforts were used to define, to improve, and to verify successes with the STS reforms which were described and published in a SUNY Monograph in 1996 (Yager, 1996) and in the current NSTA Position Statement regarding STS (NSTA, 2008-09, p. 242). The information regarding program effects on students all became important data for gaining approval by the U.S. Program Effectiveness Panel (PEP) as well as data for annual reports to NSF and summary reports for each funding period to share with all teachers, administrators, and parents.

The questionnaire consisting of personal information and the 5-choice options for definitions of the eight terms was administered to students randomly selected in homerooms by homeroom teachers in each school where teachers were willing to collaborate. Teachers for 4th graders were permitted to read each item for students when there were reading problems as has been the situation for the samples used by the National Assessment of Educational Progress (NAEP) nationally.

The eight terms selected for the study were selected from 4th grade textbooks with the following definitions:.

Volume – amount of space inside an object;

Organism – any object that is alive;

Motion – a change in position of an object;

Energy – what makes objects in a system interact;

Molecule – smallest unit of material that has the original features of the material;

Cell – small building units of living things;

Enzyme – substances which control all chemical changes in living systems;

Fossil – any evidence of past life.

With the initiation of the Iowa Chautauqua Professional Development Project in 1983 (orchestrated by NSTA and funded by NSF), five centers were typically established each year at sites across the state where enrolled teachers were invited to sample their students at the three grade levels similar to efforts by the National Assessment of Education Progress dealing with concept mastery and their first efforts with a focus in the affective domain (NAEP, 1978). Information was collected prior to direct experiences with the Chautauqua workshops over a summer -- often continuing for three consecutive academic years. Some of these studies looked at differences across grade levels as new teaching strategies were developed and used. Using social issues as organizers was criticized by science educators because of the lack of focus on basic science concepts. There was interest in seeing if such new instructional emphases resulted in less concept mastery and/or less ability to recognize accepted definitions. Of course some could (and do) criticize that the recognition of correct definitions of science terms has little to do with student understanding and ability to use the terms in other situations.

Iowa SS&C, which operated from 1990 through 1997, included all science teachers in 20 Iowa districts where the key teachers helped to convince all teachers and school administrators to be involved in such a national reform effort. Teachers and students were initially selected first from middle schools – the focus for the SS&C project. Many of the central staff members for Iowa SS&C were teachers involved during the Chautauqua Project, 1983-present. It is interesting to note that the focus on science terminology as such diminished radically over the 1985-2006 interim (Kimble, 1999). Data collection continued, however, because of the interest on the part of many teachers concerning results that were initially reported in 1985.

The Chautauqua program, the Iowa SS&C, and the continuing Title IIa workshops were all professional development projects designed to help science teachers develop their own learning processes through inquiry teaching and learning (Dass & Yager, 1999). The programs were designed for in-service science teachers in grades K - 12. However, the participants in the programs were working together in cooperative learning groups to create science inquiry activities that arose from participants' questions, curiosities, and experiences. The programs emphasized learning science content using inquiry activities that were student-centered (actually proposed, planned, and carried out by students). Moreover, they all focused on a model for inquiry-based science instruction with the teachers inquiring about their own teaching. The primary goal of the professional development was to increase the skills of in-service science teachers of science by indicating needed systemic changes in science instruction in the classrooms of all participating teachers. Basic to all the projects was the idea that teachers need to collaborate with both their students and with each other as well as with school administrators, parents, and community leaders as improvements and changes are planned. A return to a focus on major science terms seemed important to pursue where the teaching approach moved to little or no focus on such terms per se. Learning was defined more basically as evidence for successes on outcomes other than concept mastery. Instead the major focus became ability for students to use their ideas and skills which were too often merely listed in curriculum outlines, textbooks, and state standards. Often research can and has focused on much more important instructional outcomes than defining textbook terms.

This research effort included administering the instrument to students to learn of their successes with selecting correct meanings for the eight terms from randomly selected classrooms by teachers not registered for earlier P.D. Programs in Iowa. The 1985 research was replicated for each of the four years (1985, 1990, 1995, & 2000). New teachers and samples of students from at least one classroom of teachers were selected by Teacher Leaders at sites across Iowa. Table 1 indicates grade levels as well as numbers of teachers and students comprising the sample for this study. Other smaller teacher-generated studies were conducted at the school sites and in other schools in addition to those involved in this study with data collected at five year intervals between 1985 - 2000.

Table 1. Teachers and Students Involved with Choosing Correct Definitions of Science Terms over the 1985-2000 Interim

Year	Grade Level	Number of Teachers	Number of Students
1985	4	50	1480
	8	23	690
	12	21	720
1990	4	37	840
	8	21	643
	12	26	712
1995	4	33	870
	8	24	472
	12	26	486
2000	4	40	911
	8	31	842
	12	28	731

Results

The percentage of the students sampled at each grade level who were able to select the correct meanings for the eight terms is indicated in Table 2. Several findings emerge that are of interest. Perhaps most striking is the fact that there is no consistent growth in student ability to recognize correct definitions of selected terms. Eighth grade students outperformed fourth grade students on three of the terms (Volume, Molecule, and Cell). The differences (that tend to favor eighth graders) were often very small in several instances. Little increase in recognizing definitions of terms occurred for 8^{th} and 12^{th} grade students.

Table 2 indicates that 4^{th} graders performed highest on the term “organism,” and lowest regarding the term “cell” over the 1985-2000 interim. Eighth graders performed highest on the term “volume”, lowest on the term “enzyme” over the 1985-2000 interim. Table 2 also indicates that 12^{th} graders performed highest on the term “motion” in 1985, on the term “volume” in 1990, 1995, and on the term “organism” in 2000. They performed lowest on the term “enzyme” over the 15year interim.

Table II. Percentage of Students Selecting Correct Definitions for Eight Science Terms

Concept	4th Grade				8th Grade				12th Grade			
	1985	1990	1995	2000	1985	1990	1995	2000	1985	1990	1995	2000
Volume	29	28	24	29	75	71	68	70	57	63	72	61
Organism	66	65	59	53	67	65	60	63	61	54	61	66
Motion	41	43	40	36	65	60	63	61	66	58	54	59
Energy	40	38	36	30	54	52	48	52	39	52	49	56
Molecule	25	20	18	12	54	40	50	47	53	47	47	49
Cell	15	18	16	14	46	38	44	40	44	42	48	41
Enzyme	23	20	20	18	24	19	20	19	21	19	24	19
Fossil	36	35	26	30	54	48	48	42	48	46	51	52

Line graphs were developed and included as Figures I through VIII. The graphs allow the reader to note visually the results and to focus on comparisons across the three grade levels for each term. Some general observations can be made.

Figure I indicates that 8th and 12th graders outperformed 4th grade students on the term “volume”. The 8th graders showed highest performance and 4th graders resulted in lowest performance regarding the term “volume” over 1985-2000 interim. During the study, 8th graders performed highest in 1985, and 4th graders performed lowest in 1995.

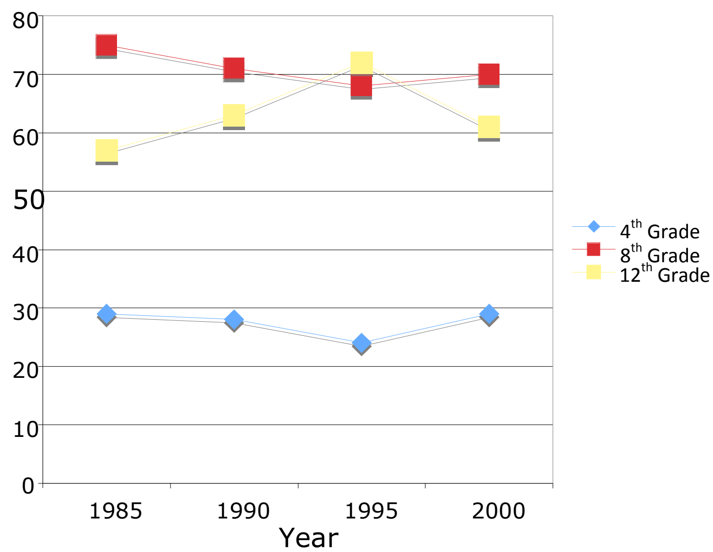


Figure I. Percentage of Students Selecting a Correct Definition for the Term "Volume"

Figure II indicates that 4th graders performed as well as 8th and 12th grade students on the term "organism". The 4th and 8th graders performed highest in 1985 and lowest in 2000. The 12th graders performed highest in 2000 and lowest in 1990 over the 1985-2000 interim.

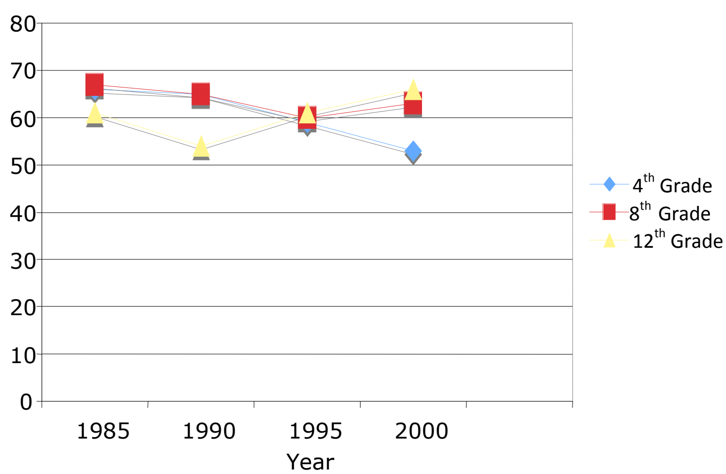


Figure II. Percentage of Students Selecting a Correct Definition for the Term "Organism"

Figure III indicates that 8th and 12th graders outperformed 4th grade students regarding the term "motion". The 4th graders performed highest in 1990 and lowest in 2000. The 8th graders performed highest in 1985 and lowest in 2000. The 12th graders performed highest in 1985 and lowest in 1995 over the 1985-2000 interim.

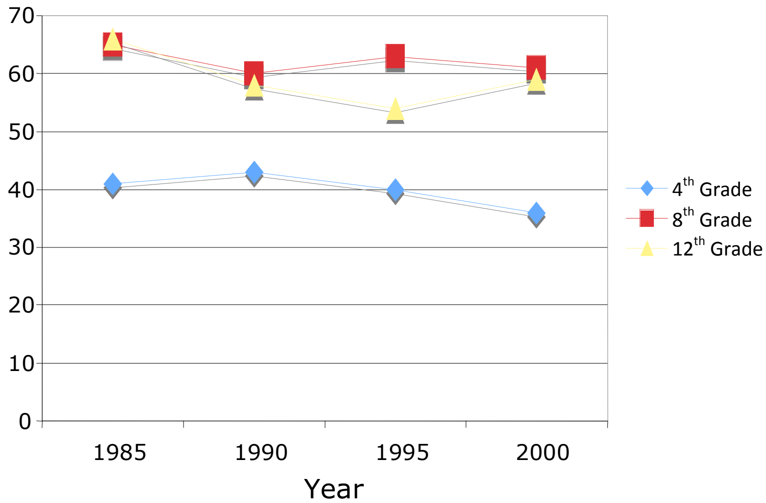


Figure III. Percentage of Students Selecting a Correct Definition for the Term "Motion"

Figure IV indicates that 8^{th} and 12^{th} graders only slightly outperformed 4^{th} grade students with respect to the term "energy". The 4^{th} graders performed highest in 1985 and lowest in 2000. The 8^{th} graders performed highest in 1985 and lowest in 1995. The 12^{th} graders performed highest in 2000 and lowest in 1985 over the 1985-2000 interim.

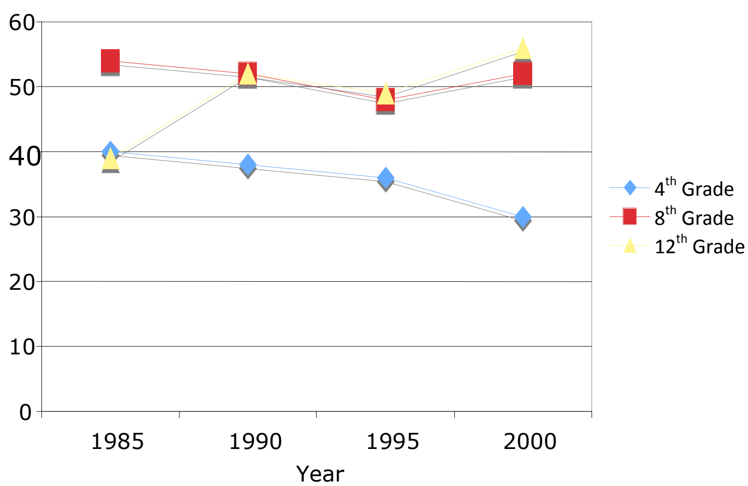


Figure IV. Percentage of Students Selecting a Correct Definition for the Term "Energy"

Figure V indicates that 8^{th} and 12^{th} graders only slightly outperformed 4^{th} grade students regarding the term "molecule". The 4^{th} graders performed highest in 1985 and lowest in 2000. The 8^{th} graders performed highest in 1985 and lowest in 1990. The 12^{th} graders performed highest in 1985, lowest in 1990 and 1995 over the 1985-2000 interim.

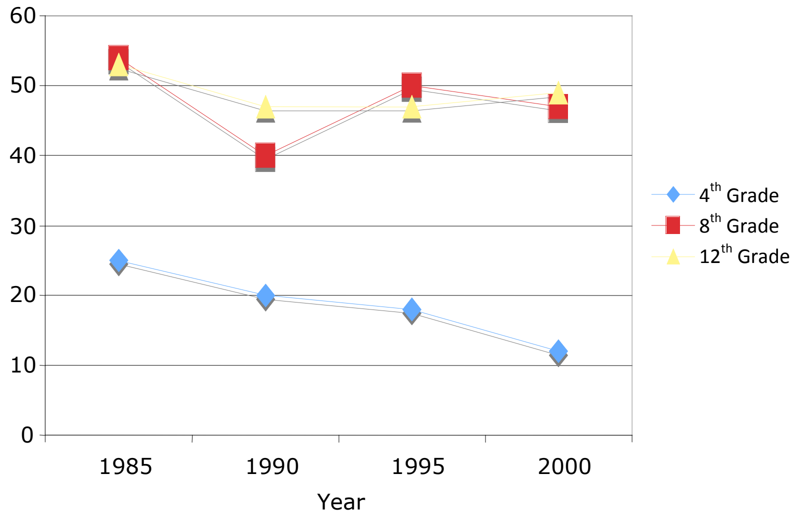


Figure V. Percentage of Students Selecting a Correct Definition for the Term "Molecule"

Figure IV indicates that 8th and 12th graders outperformed 4th grade students regarding the term "cell". The 4th graders performed highest in 1990 and lowest in 2000. The 8th graders performed highest in 1985 and lowest in 1990. The 12th graders performed highest in 1995 and lowest in 2000 over the 1985-2000 interim.

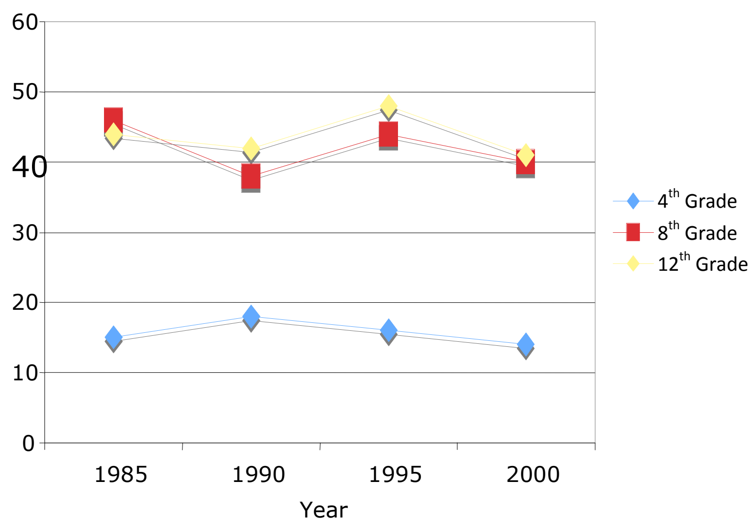


Figure VI. Percentage of Students Selecting a Correct Definition for the Term "Cell"

Figure VII indicates that 4th graders performed as well as 8th and 12th grade students regarding the term "enzyme". The 4th graders performed highest in 1985 and lowest in 2000. The 8th graders performed highest in 1985 and lowest in 1990 and 2000.

The 12^{th} graders performed highest in 1995 and lowest in 1990 and 2000 over the 1985-2000 interim.

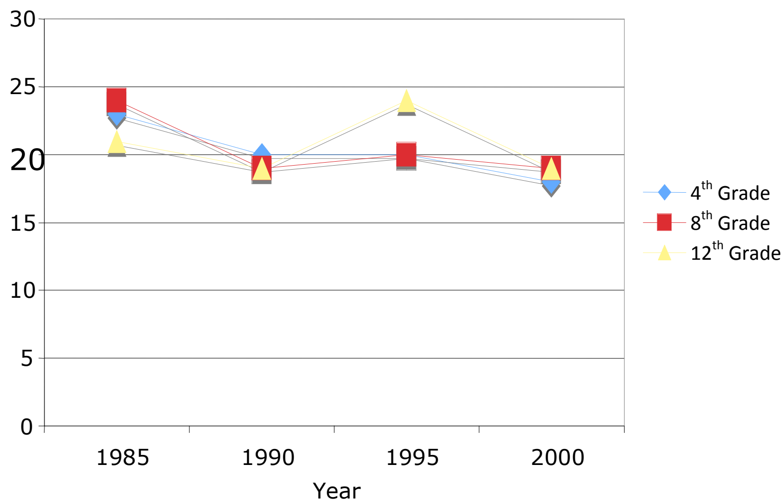


Figure VII. Percentage of Students Selecting a Correct Definition for the Term “Enzyme”

Figure VIII indicates that 4^{th} graders performed as well as 8^{th} and 12^{th} grade students regarding the term “fossil”. The 4^{th} graders performed highest in 1985 and lowest in 1995. The 8^{th} graders performed highest in 1985 and lowest in 2000. The 12^{th} graders performed highest in 2000 and lowest in 1990 over the 1985-2000 interim.

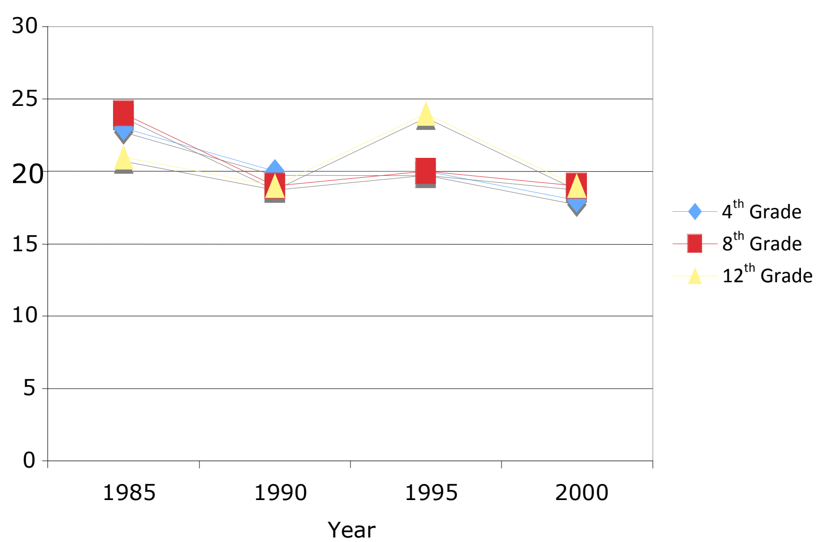


Figure VIII. Percentage of Students Selecting a Correct Definition for the Term “Fossil”

Discussion

Recent reform documents recommend new teaching and learning approaches to science education. The reason is that most teachers use the textbook as the major source for conveying information to students. While textbooks may include basic information about science subjects, they typically overemphasize vocabulary and factual information. Teachers feel pressured to make sure that students “get it all”. They often ask students to memorize specific words and facts (NRC, 1996, AAAS, 1997, p. 8). The results of this study suggest that merely recognizing accurate definitions of words and facts does not increase use of real understanding of the terms. More importantly perhaps is that a focus on local, current, and personal problems does not decrease student ability to recognize the definition of terms.

Overall, the results suggest that an emphasis upon vocabulary development is ineffective and/or misleading in terms of the recent reform objectives of the school science programs. In fact, it can be stated the school science programs for the students and schools studied seem ineffective in increasing real understanding and use of the selected science terms across the four through twelve grades. This experience was the first introduction to current efforts to change instructional focus. More information is needed concerning student ability to use the terms in completely new situations. It would also be of interest if information were not limited solely to textbook definitions and the teaching focus of the teachers as they were newly enrolled in a Chautauqua series.

The results of the study also indicate that help is needed for science teachers to realize the features recommended for improving science education that are set forth in reform documents, especially the National Science Education Standards. Teachers should be involved in helping develop strategies for promoting deeper and more meaningful understanding of science, including science terms. These strategies need to adopt new teaching and learning skills that stress science understanding rather than an over emphasis on rote memorization. They should help students move beyond the level of simple recognition of the meaning of science terms and concepts (NRC, 1996, 2007). Marzano's six points mentioned initially can help with fostering understanding and use of major terms (Marzano, 2009). But, these may not result in the teaching recommended by the NSES (p. 52).

More study is needed regarding varying contexts where teachers focus more than on the terms and their textbook meanings. Perhaps more attention to situations and reasons for use of the terms should be explored by more teachers. Interesting results and many differences among teachers (especially for the long-time teacher leaders) and new teachers involved with the professional development program could provide more insights concerning the data reported in this study. The important finding is that the perceived importance of “vocabulary first” is flawed. Perhaps meaning comes from using and developing the meaning in a variety of settings. Perhaps, too, the use of science concepts and process skills in new contexts may provide the best evidence that real learning has occurred! Certainly real world contexts make it easier to teach with the unification of process skills with science concepts as advocated in the NSES. Use can provide evidence of real learning as well as the teaching features which encourage it to occur.

Summary and Conclusions

Eight science terms were studied in terms of their stated meanings in textbooks by 4th, 8th, and 12th grade students who were randomly selected from twenty school districts in Iowa. The results indicate that the students do not increase in terms of percentages who master the definition selected for the science terms, especially with respect to the 8th and 12th grades. Regarding the terms “organism”, “enzyme” and “fossil” 4th graders performed as well as 12th grade students. Teaching for such specific and continuing mastery of terminology is questioned by the results. The results diminish the importance for all students to know and be able to use technical language developed and used for convenience among practicing scientists. Students do not increase in their abilities to define the terms over the grade levels – and perhaps more importantly, no evidence has been provided for their use in school nor in life generally – across the grade levels.

In some cases some of the terms were used in studies of what happens in the same schools at the next grade level. Further study is planned regarding issues related to student ability levels, socio-economic levels of students, and gender.

In the last 50 years, there has been little actual reform in American education. Educational policies and programs have recommended significant changes, but classroom practices have not changed. Apparently, the practice and theory of reform do not coincide (Bybee, 1993). Real reforms in science education will occur only if teachers change their ideas about the meaning of science and their views of effective teaching (Yager, 2000). When teacher beliefs are incompatible with the philosophy of science education reform, a gap develops between the intended and the implemented principles of reform (Levitt, 2002).

References

American Association for the Advancement of Science (AAAS, 1965). The psychological bases of Science—A Process Approach (SAPA). Washington, DC.

American Association for the Advancement of Science (AAAS) (1997). Science for All Children. Washington, DC: National Academy Press.

Bybee, R. W. (1993). *Reforming science education: Social perspectives and personal reflections.* New York, NY: Teachers College Press.

Dale, E. (1962). Vocabulary development and reading instruction. *Reading in modern communication.proceedings of the annual reading institute.* Philadelphia, PA: Temple University. 1, pp. 17-31.

Dass, P. M. & Yager, R. E. (1999) The Iowa Chautauqua Program: Advancing reforms in -12 science education. *Science Education International*, 10(2), 33-38.

Harms, N. C. & Yager, R. E. (1981). *What research says to the science teacher, Vol. 3.* Washington, DC, National Science Teachers Association.

Hurd, P. DeH., Robinson, J. T., McConnell, M. C., & Ross, N. M. Jr. (1981). *The status of middle and junior high school science, Volume I and summary report, Volume II technical report.* Louisville, CO: Center for Educational research and Evaluation, BSCS.

Kimble, L. L. (1999). *A comparison of observed teaching practices with teacher perceptions of their teaching during and following major funding.* Unpublished doctoral dissertation. University of Iowa, Iowa City, IA

Levitt, K. E. (2002). An analysis of elementary teachers' beliefs regarding the teaching and learning of science. *Science Education,86*,1,1-22.

Marzano, R. J. (2004). *Building background knowledger for academi achievement: Research on what works in schools*. Alexandria, VA, ASCD.

Marzano, R. J. (2009). The art and science of teaching: Six steps to better vocabulary instruction. *Educational Leadership, September 2009, pp. 84-85.*

Metz, K. E. (1995). Reassessment of developmental constraints on children's science instruction. *Review of Educational Research, 65, 93-127.*

Miller, J., Suchner, R., & Voelker, A. (1980). *Citizenship in an age of science.* Elmsford, NY: Pergamon.

National Assessment of Educational Progress, (NAEP) (1978, May). *The third assessment of science ,1976-77;* 08-S-08; released exercise set. Denver, CO.

National Research Council (NRC) (1996). *National Science Education Standards.* Washington, DC: National Academy Press.

National Research Council (NRC) (2007). *Taking science to school: Learning and teaching science in grades K-8. Washington*, DC: National Academy Press.

National Science Teacher Association (NSTA) (2008). *Science/Technology/Society: A new effort for providing appropriate science for all (Position Statement).* NSTA Handbook. pp. 242-243.

Program Effectiveness Panel (PEP) (1992). *The Iowa Chautauqua program: A model for effecting change in the teaching and learning of science in schools*. National Science Foundation, Washington, DC.

Seymour, E., & Hewitt, N. (1994). *Talking about leaving: Factors contributing to high attrition rates among science, mathematics, and engineering undergraduate majors*. Boulder: Bureau of Sociological Research, University of Colorado.

Stake, R. E., & Easley, J. (1978). *Case studies in science education, Vols. 1 and 2.* Center for Instructional Research and Curriculum Evaluation, University of Illinois at Urbana-Champaign. US Government Printing Office, Stock No. 038-000-00376-3. Washington, DC 20402.

Trumbull, D., & Kerr, P. (1993). University researchers' inchoate critiques of science teaching: implications for the content of pre-service science teacher education. *Science Education, 77,* 301-317.

Voelker, A. M. (1982). The development of an attentive public for science: implications for science teaching. *What research says to the science teacher, Vol. 4.* Washington, DC 20402.

Yager, R. E. (1983). The importance of terminology in teaching K-12 science. *Journal of Research in Science Teaching.* 20(6). pp. 577-588.

Yager, R. E. (1996). Science/technology/society as reform in science education. State University of New York Press, Albany, New York.

Yager, R. E. (2000). The history and future of science education reform. *The Clearing House,* 74,1,51-54.

Yager, R. E. & Yager, S. O. (1985). Changes in perceptions of science for third, seventh, and eleventh grade students. *Journal of Research in Science Teaching, Vol. 22, No. 4, pp.347-358.*

Identifying Valuable Components of Student Behavior: Things They Do Right When They Solve Wrong[3]

Enrique A. Coleoni
Universidad Nacional de Córdoba
IFEG - CONICET

Laura M. Buteler
Universidad Nacional de Córdoba
IFEG - CONICET

Abstract

The present study explores and characterizes some metacognitive abilities of students in an introductory university-level physics course. This characterization is done in the context of solving problems on magnetism. The study is based on a manifold view of cognition as the one in the theoretical framework proposed by Hammer, Redish and others (Hammer & Elby, 2003, Hammer et al, 2005) according to which subjects' cognition is the result of the context-sensitive activation of cognitive resources. Within this framework, metacognition is studied together with subjects' cognitive productions. Results show that students, considered novices, have a series of metacognitive abilities, from which they can construct their metacognitive expertise. This could help to better understand the process by which students do this during their learning processes.

Correspondence should be addressed to Enrique A. Coleoni, Universidad Nacional de Córdoba, at ecoleoni@famaf.unc.edu.ar, +54 351 4334051 (401).

Introduction

Problem solving is a complex cognitive task, in which metacognitive activity plays an important role. The basic function of metacognitive activity is to control and if necessary redirect the course of cognitive activity. One important finding in metacognition in problem solving is that subjects who have a better performance in the task of problem solving, also exhibit a higher degree of development in their metacognitive abilities. As an example, Gerace (2001) points out that while in novices problem solving uses almost all available mental capacities, experts are able to think about problem solving while problem solving. Howard et al (2001) examine certain metacognitive monitoring and regulatory skills in the context of solving science problems in a computer-based learning environment. These authors were able to establish that metacognitive self-regulation is a good predictor of success at problem solving. On the basis of results such as the ones mentioned, instructional environments have been

[3] Partial results of the present work were presented at the *"2007 Foundations & Frontiers in Physics Education Research Conference"*, Bar Harbor, Maine.

designed to help subjects mimic expert behavior and thus acquire certain metacognitive habits. Reports of these instructional environments show that training students in such metacognitive behavior positively affects their problem-solving performance. Georghiades (2004) introduced metacognitive activities within science instruction through classroom discussions directed at reflective thinking and diary-like notes on students´ reflections on classroom activities. The authors found that students trained with these activities retained the contents taught through a longer periods of time. Berardi-Coletta et al (1995) compare the performance on the solving of a novel problem by two groups of students. The control group received traditional instruction while in the experimental group students were either prompted to describe their actions or to give reasons for them. The purpose of asking for reasons was to bring students' own thinking into their focus of attention. The authors found that students in the experimental group performed better when solving a new problem.

The examples just shown above allow us to pinpoint two important findings of the research on metacognition in problem solving. First, as pointed out by Gerace (2001), metacognitive skills are a component of subjects' expertise. Furthermore, as illustrated by the studies of Berardi-Coletta (1995) and Georghiades (2004), when subjects are "trained" to mimic the metacognitive behavior that experts exhibit, this has a positive influence on their problem solving performance. This indicates that mimicking experts helps students to build their metacognitive expertise. These results, however, do not tell us how students do this.

On the other hand, the cognitive nature of metacognition also often stands in the way of differentiating what should be called metacognitive and not simply cognitive. As some authors point out, the already fuzzy concept of metacognition "has become even fuzzier due to a ballooning corpus of researchers of widely varying disciplines and for widely varying purposes" (Hacker, 1998, p 2).

These two difficulties: a) the fact that metacognitive abilities are known to be characteristic of experts, but it is still not clear how they are constructed and b) that the boundaries between cognitive and metacognitive phenomena is difficult to establish, could be overcome in part if cognitive and metacognitive activities could be regarded as different aspects of the same phenomenon.

Recently, Hammer and others (Hammer & Elby, 2003; Hammer et al, 2005), based on previous results of diSessa & Sherin (1998), propose a view of cognition based on what they call *cognitive resources*. In this approach, when subjects undertake a cognitive task, they activate a subset of their available cognitive resources, in a way that is context-dependent. Thus, subjects' cognitive as well as metacognitive behavior is the result of the activation of these resources. One possible (non-exhaustive) way to classify cognitive resources is to divide them into two categories: *conceptual* resources and *epistemic* resources. The first group, i.e. conceptual resources, is the set of resources that enable subjects to reason about physical situations. An example of one such resource is "*the more, the less*". Mapped onto a particular situation, such as looking at an object at a certain distance, this resource can lead subjects to reason that *the more the distance* from an object, *the less the size* of the object will appear to be. When resources are activated in situations such as these, they are said to be mapped onto the elements of that situation. In this example, the resource *the more, the less* is mapped onto the distance from the observed object and the apparent size of that object.

Epistemic resources are the ones that enable subjects to deal with their available knowledge. When facing a cognitive task, subjects pay more attention to certain traits than to others, and also adopt a particular behavior which they find (consciously or unconsciously) is an adequate response to the situation. Since the activation of epistemic resources is also context-sensitive, the same subject may exhibit different behaviors in different situations. An example Hammer et. al. (2005, p 102) offer for this is the case of a student given the name of “Louis”. This student viewed learning Physics as two completely different cognitive tasks. First he approached the task as one of memorizing “every word of the homework solutions”. After performing poorly on a midterm exam and an interview with one of his tutors, Louis decided to try one of his advisor´s suggestions and think of an analogy he would make up for a ten year old when studying Physics. He realized that he had experience working with children and was able to use the idea of making an analogy in the same way as he would explain something to a 10-year old. As a result he was able to re-structure his ideas about Physics and did significantly better on his next exam. Hammer et. al. (2005) use this case as an example to show how Louis was able to activate the resource of building knowledge from what is already known (by building an analogy) in the context of tutoring small children, and improved the way he learned Physics when he was able to activate this same resource in the context of studying for his exam. The point that is supported with this example is that the ability to use analogies is something that Louis can deploy depending on context, and therefore it is not a unitary cognitive element. Rather, it is the result of activating finer-grained cognitive elements, and this activation is context dependent. It would not be possible to understand Louis´s behavior if the ability to build analogies were considered a unitary cognitive element. This would not allow us to understand why he would not use an ability he has in a situation that calls for it.

When confronted to a cognitive task such as solving a Physics problem, students activate certain conceptual resources to reason about the physical situation, and also epistemic resources to administrate their previous knowledge. Two epistemic resources are of particular interest in the present study. They are related to different stances subjects can adopt regarding their cognitive activity:

Understanding: a student activating this resource will be satisfied with his own description of the situation at hand.

Confusion: activating this resource will allow a student to manifest dissatisfaction due to an internal incoherence between two or more of his/her ideas regarding a given situation.

These resources are closely related to students’ metacognitive activity of *checking*. As Hammer and Elby (2003) point out (cic), “epistemic resources may serve the role of helping to activate metacognitive resources; or they may turn on in response to metacognitive activity, to play an administrative role”. Once more, an appealing feature of a resources-based view of cognition (and metacognition) is that due to the contextual activation of resources, the same subject may activate a different set of (conceptual and epistemic) resources in different contexts, and thus exhibit either an expert-like or novice-like behavior. The shift from novice to expert could then be related to a higher refinement of resources, the generation of new resources, or a higher degree of adequacy in the activation of resources, which enables the subject to efficiently activate the most convenient resources in the situations in which they result fruitful.

Viewing cognition and metacognition as the result of the activation of conceptual and epistemic resources raises interest in describing, from among the epistemic resources students activate, those which enable them to perform metacognitive activities. The activation of these resources, which will be referred to as metacognitive, is what enables students to check and redirect the course of the cognitive task at hand. Therefore, we shall give the name of metacognitive resources to those epistemic resources that, when activated, allow students to monitor and/or redirect the course of their cognitive activity of problem solving. Since this activity is in turn envisaged as the activation of one or a set of conceptual resources, we shall say that the activation of metacognitive resources is what enables students to check their understanding in terms of the conceptual resources they have activated, and eventually to change their activation if necessary. We do not attempt to achieve a thorough description of the metacognitive activities novices cannot do or do incorrectly (as compared to experts). Instead, our purpose is to better comprehend the metacognitive activities they *are* able of engaging in and to be better prepared to design instruction in a way that is more efficient to promote the refinement of those abilities. In terms of metacognitive resources, the present work reports the finding of some metacognitive resources that a group of novices was found to activate during a problem solving activity.

The aim of the study is not to propose an instructional strategy aimed at fostering the mimic of novice behavior instead of expert behavior. The results that we seek are to identify certain resources that could be involved in their (still underdeveloped) metacognitive activity. This will help us understand the process by which students are able to learn, for instance, when instruction favors mimicking of expert behavior.

The Study

This exploratory study aims at the characterization of the metacognitive activity exhibited by students of an introductory university-level physics course, while solving problem situations dealing with topics of magnetism.

Participants

Students participating in the study were Chemistry majors who had recently finished this Physics course which is the second one that they take in the second semester of the first year of their career. The characterization was done on the basis of the metacognitive resources, as well as other epistemic resources, activated by nine students who volunteered to participate in the study. The instruction of these students, during the course, included the topics of forces acting on electric currents in the presence of magnetic fields. An important example of such interactions discussed during the course was that of the torque acting on closed loops of current placed in magnetic fields. In such cases, these loops of currents were described in terms of their associated magnetic moment. In the course these students had taken in the semester immediately before, they had thoroughly discussed the concept of mechanical equilibrium. Examples of mechanical systems in equilibrium included springs holding masses, masses lying on different surfaces, strings holding masses, etc.

Data Collection

Since the study is of an exploratory type, we conducted several interviews, in which the resources we wished to identify could be evidenced. The underlying assumption is that if the activation of a particular resource is evidenced, then we can assume that the activation occurred. Case studies, such as the present one, are not sufficient to support or disregard any particular hypothesis, but are valuable in providing evidence for the existence of elements such as metacognitive resources. The analysis was done following the idea that the activation of epistemic resources can help in the activation of metacognitive resources, and that epistemic resources can in turn be activated in response to metacognitive activity, as already pointed out by Hammer et al (2003). This circular relation between the activation of metacognitive resources and other epistemic resources, and the fact that both epistemic as well as conceptual resources are activated during problem solving calls for the observation of these activations as part of the same phenomenon, and therefore the activation of one type of resource is reported in the context of the others. Another study has been carried out in which the focus was directed at the characteristics of the activation of conceptual resources in a similar setting (Buteler & Coleoni, 2010). As for the methodology used to obtain data, two characteristics of the interviews were relevant:

1. Subjects were interviewed in groups of two (in one case three), in order to favor the flow of verbalizations.

2. Problem statements were not presented at once, as in a printed sheet of paper, but in a sequential manner, sentence by sentence. This also had the purpose of increasing the flow of verbalizations, since students had more time to produce them, and they could be allocated more specifically to each portion of the problem. This way of collecting data sentence by sentence was previously used in another study (Buteler & Coleoni, 2006). Interviewers' participation was limited to keep the flow of students' verbalizations, and occasionally to require clarification, without referring to the correctness of students' productions.

The purpose of the present report is to characterize metacognitive activity in terms of the activation of metacognitive resources. The main assumption sustaining the way data are analyzed is that the activation of certain epistemic resources serves as the basis for the activation of metacognitive resources, which are the ones that enable subjects to carry out metacognitive activities. As a result of this metacognitive activity, different resources (epistemic and conceptual) can in turn be activated. Therefore, analyzing the activation of metacognitive resources requires also reporting the activation of other, conceptual as well as epistemic, resources.

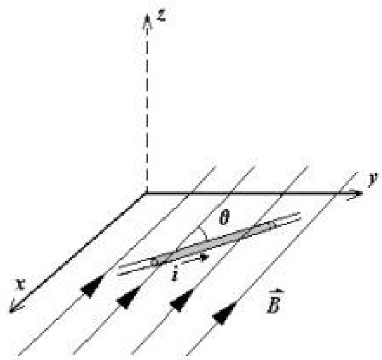
The analysis of data is done on the basis of a case study. We seek to identify the existence of such resources, and to observe their activation in the context of a problem solving situation, i.e. together with the activation of other epistemic as well as conceptual resources.

The problems used

The problems presented to students are shown in Figure I.

Problem A

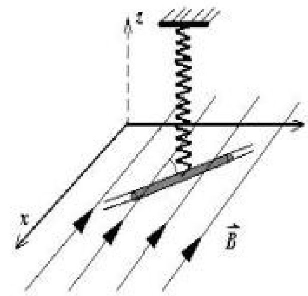
A conducting rod of length l and mass m is placed horizontally in a zone where there is a uniform magnetic field B, which is also in the horizonal plane. An electric current i passes through it. The rod forms an angle θ with respect to the direction of the field, as shown in the figure



Considering that: i=0.01 A, B=0.3 T, l=0.5 m, g=10 m/s^2 and m=0.045 kg, what is the value of θ for which the rod will be in equilibrium?

Problem B

A conducting rod of length l and mass m hangs from a spring of elastic constant k, as shown in the figure. The space in which the rod is located is affected by a constant and uniform magnetic field B. Both the rod and the field are on the horizontal plane.



1. Being k=45 N/m and m=0.045 kg, How much is the increase in the length of the spring when the rod hangs from it?
2. If an electric current passes through the rod, what will happen to the spring, compared to the situation described above?

Figure I. The problems used in the study

Results

In what follows, excerpts from the transcriptions of students' protocols are presented, to show certain metacognitive resources. These are recognized when activities are identified that enable the student(s) to change the course of their cognitive action, or in other words, to produce changes in the activation of the conceptual resources. The conceptual resources that will be shown in the transcripts are two: balance and alignment:

Balance: activation of this resource allows students to balance the effects of two opposing agents. The activation of this resource is useful to address problems in which one or more agents exert forces on an object in equilibrium.

Alignment: this resource is useful to reason about two entities that rotate in order to align with one another. An example of a fruitful activation of this resource is when it enables a person to understand the alignment of a compass needle with the existing magnetic field; or the alignment of an electric dipole with an external electric field.

As for the metacognitive resources that will be reported, and which are the focus of this study, the first of them has been named *reconciling*, has already been found to be activated in children by Lising & Elby (2005). The other resource found has been given the name "what-happens-if". A definition for both these resources is offered next.

Reconciling. When this resource is activated, the consequences from two different lines of reasoning are reconciled into one coherent description. One possibility, for example is to reconcile the reasoning stemming from everyday experience with the one generated through formal knowledge. Therefore, its activation allows students to check for coherence between available knowledge from different sources.

What-happens-if. By activating this resource, subjects evaluate their comprehension by posing this question and evaluating their own responses. The particular trait of this resource is that that the inference subjects can make by asking themselves this question is not suggested in the situation to be solved.

The activation of these metacognitive resources is seen together with the activation of the epistemic resource of *confusion*. The excerpts presented in this section serve as examples to illustrate this. A summary of the metacognitive resources reported is presented in Table I. The other cognitive resources (conceptual as well as epistemic) that are present in the transcripts are also presented in the same summary.

Table I
Summary of the resources described in the transcripts

	Resource	Description
Conceptual	Balance	activation of this resource allows students to balance the effects of two opposing agents
	Alignment	useful to reason about two entities that rotate in order to align with one another
Epistemic	Understanding	a student activating this resource will be satisfied with his own description of the situation at hand
	Confusion	activating this resource will allow a student to manifest a dissatisfaction due to an internal incoherence between two or more ideas regarding a given situation
Metacognitive	Reconciling	by activating this resource is activated, the consequences from two different lines of reasoning are reconciled into one coherent description
	What happens if	activation of this resource allows students to evaluate their comprehension by posing this question and consider their own responses

The first excerpt presented, corresponding to Ana and Guillermo, is an example showing the activation of *confusion* together with the metacognitive resource *what happens if.* Excerpts 2 (students Claudia and Pablo) and 3 (Valeria, Darío and Gustavo) show the activation of *confusion*, together with both metacognitive resources, *reconciling* and *what happens if*.

Excerpt 1: Both Confusion and what happens if are exhibited

Ana and Guillermo: (while reading problem A)

Interviewer: So, you're saying that to be in equilibrium, it has to be in the direction of the field?...

Guillermo: yes, because then... the force is times the...

Ana: 'cause then there won't be any force

*Guillermo: ...times the sine of theta (*and at the same time)

Ana: ...but... did we say everything wrong?

Interviewer: ...so... that's what you think will maintain the rod in equilibrium...

Guillermo: yes...

Ana: oh!...there's still gravity...

Guillermo: oh, ok... I don't know, 'cause I never saw gravity in these kind of cases

Interviewer: no? What do you mean?

Guillermo: there's a force pulling it down

Interviewer: does the rod have weight?

Ana: sure! (at the same time) *Guillermo: yes!*

Ana: oh! So... what? When the field aligns it… it just… falls down?!

The transcript above shows students activating *alignment* and stating that equilibrium will be reached if θ=0. Ana activates *confusion* (she is not certain of her conclusions: "did we say everything wrong?"). Also, it is possible to see her activating *what happens if*, after noticing that the rod has a finite mass and hence weight. She analyzes what would happen if the rod were allowed to rotate, and notices that it would align with the field, and then fall as a consequence of its weight. This leads her to once again activate *confusion* ("it falls down?!") This metacognitive resource is her response to her state of confusion.

Guillermo: let me see... wait... oh... so, we need a force opposite to that, pointing up 'cause... it has to be aligned with (in the direction of) the weight

Ana: 'cause actually...wouldn't it have to be in equilibrium there? With that angle, in that position, the force pointing up, I mean the force from the field, is the same as the weight… I mean, that's what we have to compute

Guillermo: yeah, we have to see if that's equal to the force from gravity

Interviewer: do you want to do some kind of computation, drawing?

Guillermo: we get 0.45 N for the weight...

Ana: (makes the computation) well, that's the magnitude

Guillermo: yes, the direction is vertical, and pointing down… in the negative z direction

Ana: so this force, magnetic force, has to have the same modulus

Guillermo: but upward... that is, vertical and... pointing up…

Ana: *and the formula for the force was....*

Guillermo: *i times B times l times the sine of the angle… so (they solve for the sine of theta)*

In this part of the transcript, Ana and Guillermo activate *balance*. Even though the complete protocol is not reported, they are satisfied with their solution, and feel they understand the situation, which is interpreted as the activation of the epistemic resource of *understanding*.

Excerpts 2 and 3: the epistemic resource of Confusion and the metacognitive resources reconciling & what happens if

Excerpt 2: Claudia and Pablo (While solving problem B)

Claudia: *(puzzled) mass m… hanging from spring?! (halts)*

Interviewer: *Anything else?*

Claudia: *we never saw anything like this…I mean… that's the first thing we… you look at the drawing and if its something we never saw we go "wow! What's this?!" if its too different, I kind of get scared…*

…

Claudia: *...I'm not so sure about this... I'd have to think some more... hmm, no, I mean, I need to make some computation to... to decide if things happen the way I think they do... 'cause the force on it will pull it up...*

Interviewer: *what force?*

Claudia: *the force of the magnetic field... upward... yes... and all the time, 'cause the field is uniform and constant…*

Pablo: *the current that is passing through there, is it going that way? … right… doesn't say anything, so… what if the current were going that way?… (hands gesturing the right hand rule) it would be pulled down… ok, so we are actually assuming that the current is like in the first problem… but it doesn't really say anything about it.*

These students had carried out a physically correct solution in problem A. When addressing problem B, they make a qualitative analysis of the problem. Having activated the resource of *confusion*, Claudia makes an attempt to *reconcile* her ideas of *balance* with a formal expression, in order to be sure of her assertions (although she does not manifest being confused, she does express a strong uncertainty regarding her ideas). The resource *what happens if* allows Pablo to monitor and refine his understanding in deciding whether the rod tends to stretch the spring further or not.

Except 3: Valeria, Darío and Gustavo (While solving problem B, item 1)

Valeria: it's the same problem... only with that little spring there... well, maybe the angle isn't the same, but its the same problem…

Gustavo/Darío: but it doesn't say anywhere that there is a current through the rod...

...

Interviewer: what's gonna happen there?

Valeria: nothing…

Darío: if there's no current, there is no magnetic moment, and there is no torque

Valeria: and the rod is just gonna stay there, as it is

Interviewer: and the spring… why is it there?

Valeria: just to make things more complicated! (laughing)

Gustavo: (reads first question)

Valeria: what was the formula like?

Gustavo: yeah, for the spring... –k times the distance...

Valeria: oh, yes, times the "stretching"...

Darío: it's the force opposite to the weight...

Valeria: what?!

Darío: thing is... I'm not sure if what I'm saying is right...

Gustavo: yeah but... the field does have to do something on the rod, right? I mean, you don't need a current... if you have something metallic, you put it near a magnet, there is an attraction

Valeria: but... how do you mean? (To Gustavo)

Gustavo: if you put something metal near a magnet, the magnet's gonna attract the metal "thing"

Valeria: if you have a current...

Gustavo: *no! In the fridge door there is no current, and the magnets stick to it*

Darío and Gustavo realize that there is no current passing through the rod (which is an essential feature based on which *alignment* was activated in the previous problem). Also, it is possible to observe Darío activating the conceptual resource of *balance*. This seems to foster the activation of *confusion* in Valeria, and an attempt from Gustavo to *reconcile* this formal description with his everyday experience. Gustavo attempts to *reconcile* two ideas. On the one hand, what they are elaborating from formal elements, and, on the other hand, the activation of a conceptual resource of *attraction* probably influenced by his everyday knowledge that "refrigerator magnets stick to metal doors"[4]

(Darío seems to have something to say, but is hesitant)

Interviewer: *Darío, what are you thinking? Tell us...*

Darío: *if it says there that the spring is making a force... the force the spring does on the rod is the inverse (for opposite) to the force the Earth does on the rod... just imagine you're hanging the rod on the spring... for it to be in equilibrium, there has to be a force from the spring equal and opposite to... I don't know... that, I know is right, I'm just not sure if it has to do with all this…*

Darío activates a control resource that consists of thinking of a problem similar to the present one. In the context of this analysis, this has been classified as *what happens if.* Thus, he poses the idea of what would happen if one would simply hang a mass/rod from a spring. The fact that he adds "I'm not sure if it has to do with all this" is indicating that the activation of these resources occur together with the activation of *confusion*.

Valeria: *(to Gustavo) why do you say that the field attracts the conducting rod?*

Gustavo: *I may be wrong, but I think magnetic fields attract metals, metals are attracted by magnets, so then the rod would tend to go that way and the spring will have to stretch more*

(What he means, as suggested by his speech and his gestures, is that the rod will be attracted in the direction of the field, and therefore will tend to move in that direction, so the bottom end of the spring will feel a force in the negative x direction. Since the upper end of the spring is fixed, the spring will be further stretched)

[4] This is further clarified by Gustavo, who explains this idea to Valeria.

Valeria: *and why couldn't it be stretched the other way? (in the positive x direction)*

Gustavo: *well, I'm not sure if it goes with or against the field, but its one of those two possibilities...*

Gustavo keeps trying to reconcile with his experience on magnets clinging to refrigerator doors, and Valeria tries to follow his reasoning. Darío attempts to reconcile his classmates' explanations (basically Gustavo's) with his formal knowledge. Since Gustavo claims that the rod will feel a force in the direction of the field, but cannot decide whether it will be in the positive or negative x direction, Darío then tries to reconcile this ideas of *attraction* with the formal knowledge that magnetic fields are generated by permanent magnets or by currents (via magnetic dipolar moments) He therefore tries to imagine the orientation of the magnet equivalent to the rod, but as there is no current, there is no magnet associated to it.

Darío: *sure, it depends on the field the conductor makes... for that you need to consider the conductor as a magnet too, and see if that magnet will be attracted or repelled by the field... but since there is no current, to me there is no magnet.*

Darío attempts to reconcile his classmates' explanations (basically Gustavo's) with his formal knowledge. Since Gustavo claims that the rod will feel a force in the direction of the field, but cannot decide whether it will be in the positive or negative x direction, Darío then tries to reconcile this ideas of *attraction* with the formal knowledge that magnetic fields are generated by permanent magnets or by currents (via magnetic dipolar moments). He therefore tries to imagine the orientation of the magnet equivalent to the rod, but as there is no current, there is no magnet associated to it.

Discussion

This study shows students´ activation of certain metacognitive resources namely those named as *reconciling*, and *what-happens-if*. These activations occurred together with the activation of the epistemic resource of *confusion*. It was also possible to observe how the activations of different resources are related to each other, and that the activation of the mentioned metacognitive resources can lead to changes in the activation of different conceptual resources.

Also, it was observed that the effect of activating metacognitive resources is not always that of redirecting cognitive activity towards formally "correct" results. Such is the case of Valeria and Gustavo, when after activating the resource of reconciling, keep the activation of alignment or that in any case, an attraction in the direction of the field will be added to the alignment of the rod.

Previous work on metacognition has described certain metacognitive abilities of experts, and that they are related to a good level of problem solving performance. These findings have led to the design of instructional strategies aimed at fostering students' expertise in these abilities. This is often achieved by inducing students to mimic expert behavior. In the present study we intend to make a step forward in understanding *why* this is often successful. That is, we aim at better understanding the process by which students build their metacognitive expertise during such activities. Results from the present study show that students (novices) have metacognitive resources available such as *reconciling* and *what happens if*. These are the basis on which they can build their metacognitive expertise. These findings, along with the fact that other metacognitive resources can be present and described in the future, could allow for a more effective way of fostering their improvement through instruction.

Although the activation of metacognitive resources is seen to occur together with the activation of the resource of *confusion*, some considerations are in order regarding this apparent co-activation of resources. In the case of Valeria, for example, *confusion* does not seem as effective to promote a change in the activation of *alignment* and the activation of *balance*, as is seen with Darío (who activates *balance*) and Gustavo (who activates *attraction*). Also, it is Darío and Gustavo who bring new considerations into the solving process the three are carrying out together. Although they all activate some metacognitive resource when they have activated the resource of *confusion*, they do not do this at the same time. Valeria, for example, maintains confusion much longer than her peers (Darío and Gustavo). Therefore, the question arises of whether *any degree of confusion* is as useful a trigger for a student to activate metacognitive resources.

Another important issue about analyzing students' cognitive and metacognitive productions in terms of the cognitive resources they activate lies in the fact that the productions of students that are usually regarded as mistakes contain valuable information of their potential abilities. Moreover, those resources are the tools they already use to address problem solving, and therefore a sensible instructional decision would be to improve our understanding of those resources and the details concerning their activation. In this respect, further study of students' metacognitive resources should involve situations with the potentiality of generating different degrees of confusion, and analyzing the conditions under which the activation of confusion is favorable for the activation of metacognitive resources. Therefore, analyzing situations in which students are confused should be considered a potentially useful task, just as analyzing the mistakes they make.

References

Berardi-Coletta, B., Buyer L.S., Dominowsky, R.L., & Rellinger, E.A. (1995). Metacognition and problem-solving: a process-oriented approach. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21, 205-223.

Buteler L.M. & Coleoni, E.A. (2010) Solving a problem with the aid of another: A teaching strategy in electromagnetism. *Journal of Science Education*. ISSN: 0124-5481. In Press (To appear in vol. **11**, # 1, 2010)

Buteler, L.M. & Coleoni, E.A. (2006) Los recursos cognitivos, una perspectiva útil para estudiar la resolución de problemas en física. Presentation at the *IV Congreso Iberoamericano de Educación Científica*. Organized by Concytec. Lima, Perú. November 7-9

diSessa, A. & Sherin, B. (1998). What changes in conceptual change? *International Journal of Science Education, 20 (10),* 1155-1191.

Georghiades, P. (2004) Making pupils' conceptions of electricity more durable by means of situated metacognition. *International Journal of Science Education*, *26* (1) 85-99.

Gerace, W. (2001) Problem solving and conceptual understanding. In S. Franklin, J. Marx, K. Cummings, (Eds*.) Proceedings of the 2001 Physics Education Research Conference* PERC Publishing, Rochester, NY. 33-36.

Hacker, D. J. (1998). Definitions and empirical foundations. In Hacker, D. J., Dunlosky, J. y Graesser, A. C. (Eds). *Metacognition in educational theory and practice*. 1-25. Lawrence Erlbaum Associates, NJ.

Hammer, D. & Elby, A. (2003). Tapping epistemological resources for learning Physics. *Journal of the Learning Sciences*, *12*, 53 -81.

Hammer, D., Elby, A., Scherr, R. & Redish, E. (2005). Resources, framing and transfer. In Mestre, J. (Ed). *Transfer of Learning from a Modern Multidisciplinary Perspective*. Greenwich, Connecticut: Information Age Publishing

Howard, B.C., McGee, S., Shia & R., Hong, N.S. (2001). The influence of metacognitive self-regulation and ability levels on problem solving. Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA. Retrieved August 3rd 2007 from Educational Resources Information Center.

Lising, L. & Elby, A. (2005). The impact of epistemology on learning: A case study from introductory physics. *American Journal of Physics, Physics Education Research Section, 73* (4) 372-382.

Is There Something Useful In Students' Mistakes? : A Cognitive Resources-Based Approach [5]

Laura M. Buteler
Universidad Nacional de Córdoba
IFEG - CONICET

Enrique A. Coleoni
Universidad Nacional de Córdoba
IFEG - CONICET

Abstract

The study reported analyzes the mistakes made by university physics students when solving two problems on geometrical optics and two on magnetism. It also offers other teaching contexts in which the same reasoning leading to these mistakes could lead to correct answers. Instructional implications are discussed on the basis of the results. The study is carried out using the concept of cognitive resources proposed by Redish (2004), Hammer, Elby, Scherr & Redish (2005), and Hammer (2004) in their theoretical framework. Results show that this construct is useful to characterize different kinds of "mistakes" made by students, and also that these mistakes can be regarded as a means of probing what students *do* know which in turn can be used to direct the design of useful learning environments.

Correspondence should be addressed to Laura M. Buteler, Universidad Nacional de Córdoba, at lbuteler@famaf.unc.edu.ar, +54 351 4334051 (401).

Introduction

Two kinds of approaches can be found to support strategies proposed to teach physics problem solving. One of the approaches, arising in the late 70's, is theoretically based on Cognitive Psychology and is represented by studies of "expert-novice" differences (Maloney, 1994). The other source, based on Scientific Epistemology has its main referent (at least within the Spanish-speaking community) in the "Model for problem solving according to scientific methodology", developed by Gil Pérez and collaborators, in the early 80's (Gil Perez & Martínez Torregosa, 1983).

[5] Preliminar results were presented at IV Congreso Iberoamericano de Educación Científica, Lima, Perú, 2006.

Studies on expert-novice differences are based on the distinct characteristics observed between these two groups of problem solvers (usually, experts and novices are represented by physics teachers and students, respectively). The differences reported are basically related to subjects' knowledge structure on a particular domain in physics and to the strategies these subjects use to address problems (Chi, Feltovich & Glaser, 1981, Chi, Feltovich & Rees, 1982, de Larkin, McDermott, Simon & Simon, 1980, Maloney, 1994).

The teaching strategies proposed within the expert-novice approach aim at fostering the development of expert-like behavior. They emphasize the results desired rather than students' previous knowledge. They focus on the expert-like behavior desired in students and not on the cognitive process by means of which students can build this behavior. (Foster, 2000, Heller & Heller, 1995, Huffman, 1994, Leonard, Gerace & Dufresne, 2002, Mestre, Dufresne, Gerace, Hardiman, & Tonger, 1993). Although these strategies differ from each other, they share one common trait which is to generate constraints that lead students to mimic expert behavior.

The Model for problem solving according to scientific methodology (Gil Pérez & Martínez Torregosa, 1983, Gil Pérez, 2003) proposes a way of teaching to solve physics problems based on the (simplified) characteristics of the way in which the scientific community produces and validates knowledge. The method proposes a parallelism between the student and a novel researcher, between the teacher and an experimented researcher, and between the classroom and the scientific community. The usefulness of traditional end-of-chapter problems is dismissed in light of this parallelism, since these are considered to favor a methodology of superficiality (rote application of formulas). At the same time, Gil Pérez & Martínez Torregosa (1983) and Gil Pérez (2003) propose teaching strategies which include qualitative analysis, hypothesis formulation, solution planning, analysis of partial results, thus mimicking the behavior of scientists solving actual research problems. The appeal of this proposal lies in the fact that it opens a different perspective on the problem solving task, as compared to the way it has been traditionally addressed in educational environments. Nevertheless, as will be pointed out shortly, it still dismisses what students already know.

Gil Pérez´s approach has produced various studies on scientific (and particularly physics) problem solving, many of which aim at pointing out how much students´ behavior differs from that of experts when they are instructed using traditional problems. (Becerra Labra, Gras-Martí & Martínez-Torregrosa, 2004, Guisasola, Furió, Ceberio & Zubimendi, 2003). Their results show the inappropriate cognitive habits and usual procedures of students which, according to the authors, would be reverted if students were taught following the guidelines deriving from the problem solving model proposed by Gil Pérez & Martínez Torregosa (1983). Once more, the actual starting point is not what students know or know how to do (this is characterized as inappropriate knowledge), but rather the knowledge or capacities teachers want their students to have.

Beyond the many differences between Gil Pérez approach and expert-novice differences approach, the point we wish to highlight is that the proposed teaching strategies underestimate the relevance of what students already know and/or are capable of doing for future learning. Many of these proposals are prescriptive. The basic idea underlying them is "what does a teacher or researcher consider that a student should do in order to solve problems as similarly as possible to the way a teacher or researcher does, *regardless* of what the student already knows and/or is capable of doing". Even though some of these teaching strategies have aimed at promoting a cognitive conflict -between the student's ideas and those of the teacher's or a textbook- to challenge students' previous conceptions, this has shown to be insufficient to promote a dynamical construction of knowledge (Chi, 2005, Pozo, 1996, Redish, 2004). These conflicts often end up being a unsurpassable barrier between what students really believe and the "correct" answer they must provide to their teachers.

In short, physics problem solving is in itself an activity by means of which learning can take place. This learning occurs on the basis of what students already know and are already capable of doing. It is thus quite difficult to explain how students could learn to solve problems in an expert-like fashion from the starting point of knowledge that is either wrong or inexistent.

Redish (2004), Hammer (2004) and Hammer et. al (2005), partly based on the work of diSessa (1993) and diSessa & Sherin (1998), propose a theoretical frame that allows one to approach problem solving from a perspective based on what students *do* know. Instead of focusing on the flaws of students' previous knowledge (their misconceptions), Hammer et. al (2005), propose to favor learning from the cognitive resources that students do possess, and to take advantage of them during this process. The following section presents the basic ideas of these authors. They will be used in the present study to analyze the verbal protocols of introductory (algebra-based) physics students at the university level, solving two geometrical optics and two electromagnetism (E&M) problems..

Theoretical frame

The resources framework is a framework still under development. The framework was first presented as such by Redish (2004), Hammer (2004) and Hammer et. al (2005). The development of the framework has continued since then, as accounted for by further publications (for example Tuminaro, 2004, Tuminaro & Redish, 2007, Russ, Scherr, Hammer & Mikesa, 2008, Bing & Redish, in press). Nevertheless, within the present study the concepts of the framework that will be central are those published in Hammer et. al (2005) since 1) these concepts are sufficient for the analysis of the data and 2) these concepts have not changed since the work of 2005.

Hammer et. al. (2005) propose that people possess a collection of *cognitive resources* which they activate contextually when confronted to a cognitive task. Thus, reasoning about any particular situation involves tacitly or explicitly selecting a subset from a collection of available resources. All resources are useful in some context, otherwise they would not exist. In any case, resources can be either fruitful or not to address a given situation. This means that the activation of a particular resource in a given situation can lead to either a physically correct or incorrect statement. Hammer et. al (2005) consider *conceptual* and *epistemological* resources.

Conceptual resources are those that enable people to reason about physical situations. Although they are not themselves “wrong” or “right”, they can be mapped on the particular situation in a way that can lead to “correct” or “incorrect” physical statements. From this perspective, a physically “wrong” answer could arise from a cognitive resource that in another context, or mapped in another way, can give rise to a correct statement.

Hammer et al. (2005, p. 95) pose an illustrative example that shows the usefulness of conceptual resources to understand students’ reasoning (cited from diSessa, 1993). In tests to probe conceptual understanding in physics it is common to ask students about the forces acting on a body thrown vertically upwards. Many answer that there are two forces involved: the weight, pointing down, and another force that points up which decreases as the object reaches its highest position. When asked explicitly about the forces in this highest point, they answer that the downward and upward forces are equal. In order to explain students’ response, the authors interpret that two different conceptual resources are activated. The first, called *maintaining agency*, is the need for an agent to persist in order for the corresponding effect to be observed. In this case, the agent must continue to act for the body to keep moving upwards. When asked about forces, students map *agent* on *force*. However, when thinking about the highest point in the trajectory, the same students activate *balancing* (something directed upwards that must be balanced by something directed downwards). Asked about forces, they answer that it is the upward and downward forces that must be equal. This example is illustrative of how a resources-based approach naturally fits the description of students (context-sensitive) reasoning, and provides a more fruitful theoretical tool than the “movement requires force” misconception. Within the resources framework, a conception is built when, “with reuse, a set of activations can become established to the point that it becomes a kind of cognitive unit, and so a kind of resource in its own right. For instance, an infant comes to think about “objects” in a fairly consistent way across a wide range of situations. The cognitive unit can have its own activation conditions, passive or deliberate. But once activated, the internal coherence in the resource activations is automatic… Its activation continues to depend on context, like any other resource, but its stability does not” (Hammer et. al, 2005, p. 110)

Epistemological cognitive resources operate on people’s prior knowledge and allow them to understand sources of knowledge, forms of knowledge and stances toward knowledge. Epistemological resources tend to become activate in locally coherent sets. This locally coherent set is called a *frame*. In terms of Hammer et. al. (2005): “By a frame we mean, phenomenologically, a set of expectations an individual has about the situation in which she finds herself that affected what she notices and how she thinks to act. An individual´s or group's framing of a situation can have many aspects, including social (Whom do I expect to interact with here and how?), affective (How do I expect to feel about it?), epistemological (What do I expect to use to answer questions and build new knowledge?), and others” (p. 98).

This approach implies a shift in the way problem solving is investigated. This shift goes from a researcher-centered view: which are the relevant factors for physics problem solving, as regarded by an expert, to a subject-centered view: what is it that really occurs when students solve problems and how can we take advantage of that instruction-wise. From this viewpoint, the mistakes students make during problem solving are not mere samples of “incorrect” knowledge, but rather they are envisioned as the result of the activation of their available resources. Therefore, studying these mistakes could give useful information on the productive aspects of their knowledge and on this basis think of possible instructional strategies to take advantage of those aspects.

The goal of the present study is to classify the mistakes made by 8 students on the basis of the idea of cognitive resources. These students pertain to an introductory university-level physics course and the problems they solve are two of geometrical optics and two of E&M. The classification obtained is used to predict contexts in which these mistakes would not occur. The consequences of these results are discussed regarding possible implications for instructional decisions.

The study

The present is an exploratory study which consists of the interpretation of a few cases. For this reason, transcripts of pieces of the studied protocols are presented as a substantial part of the analysis. The verbalizations for the 8 students solving the task are analyzed following the tradition of case study methodology from qualitative research. The idea is to analyze a small number of students' verbalizations to develop case studies: rich, detailed descriptions of student reasoning in each episode. Although the limited size of the sample only allows to make conclusions regarding the subjects interviewed, they are helpful to improve our understanding of the knowledge that students make use of while solving physics problems, and possible ways to take advantage of it, instruction-wise. The selected episodes are the result of a negotiation between independent interpretations carried out by three researchers (two of them are the authors of this study). The interpretations were about when and where a particular cognitive resource is activated. Given the interpretive nature of the study, the characteristics of the participants´ instruction are provided.

Characteristics of the subjects involved.

The 8 students who participated in the study were freshmen who had just finished the second introductory physics course. They were familiar with basic algebra as well as calculus knowledge, which they had covered in two courses that same year. However, this second introductory Physics course is mainly algebra based. They are students of different careers such as Pharmacy, Biochemistry and Chemistry. The institution is a public university in Argentina and, at the time of the interviews, the students had passed the course with a score of 80% or more, as marked by the school’s regulations. Students volunteered to participate in the study. The course which the students had taken covers contents of geometrical and physical optics, electrostatics, electrodynamics, magnetism, and electromagnetism. The students took two 1.5 hour lectures and two 1.5 hour problem-solving sessions every week during 15 weeks.

During the problem solving sessions on the topic of geometrical optics, all students in the course solved typical end-of-chapter problems involving reflection and refraction, as those found in introductory physics texts. Regarding mirrors, the students usually found problems in which they had to obtain size and position of images of objects placed in front of plane and spherical mirrors, and also to determine the zones in space from which an observer could completely or partially visualize those images. The students spent a total of 4 sessions (6 hr) working on such problems.

During the problem solving sessions on the topic of magnetic forces generated by currents and the Law of Faraday-Lenz, all students in the course also solved typical end-of-chapter problems as those found in introductory physics texts. Frequently, problems requested the calculation of the magnetic field generated by currents in the form of straight lines, coils and solenoids, forces exerted by external fields on conductors carrying currents; values of the total magnetic field on such situations, and the calculation of the magnetic moment of coils with current and the torque on those coils, when placed in an external field. Regarding the Faraday-Lenz law, students calculated electromotive forces generated in coils and solenoids due to variations of magnetic flux and identified the currents thus induced in these conductors. Students spent a total of 4 sessions working on such problems.

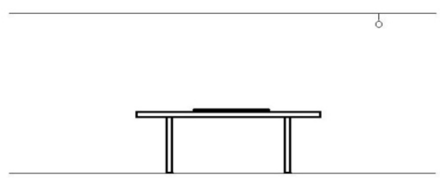
The task

Each of the participants was individually interviewed by the authors during approximately 40 minutes. Students were asked to think aloud as they read each of the sentences in the problem statements (shown in Figure I). Statement sentences subsequently appeared on a computer screen at the students' command (not all together, as in a printed sheet), allowing them to think aloud after each sentence. This technique increased the amount of verbalizations (usually quite scarce in students) and also allowed us to allocate the activation of resources to the different stages of the problem. Interviewers intervened only to ask questions when clarification was needed.

Problem 1

S1. There is a table on which a small plane mirror has been placed.

S2. The lamp lighting the room is on the roof (you may consider the lamp a point)



S3. Mark the boundaries of a region from which a person could observe the image of that lamp.

Problem 2

S1. Consider a person standing in front of a wall on which a plane mirror is to be hung.

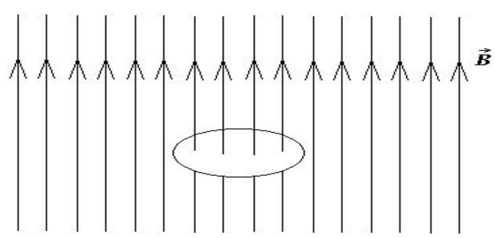
S2. This person is 1.65 m tall and his eyes are 1,55 m above his feet

S3. Compute the height at which the mirror should be hung and the minimum height it should have in order for the person to see his complete image on it.

Problem 3

S1. A conducting coil of area A and resistance R is placed in a region of space where there is a uniform magnetic field B.

S2. The plane of the coil forms a right angle with the direction of the magnetic field.



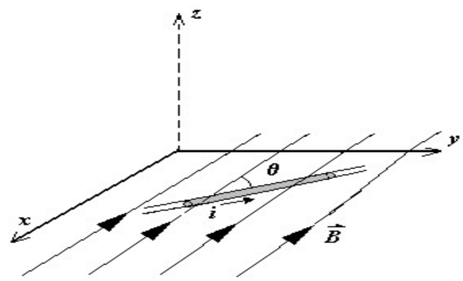
S3. The intensity of the B field raises at a rate of 0.1 Tesla per second.

S4. Compute the intensity of the current induced in the coil, knowing that its area is $A = 0.01\ m^2$ and its resistance is $R = 10\ \Omega$

Problem 4

S1. A conducting rod of length ***l*** and mass ***m*** carries an electric current of intensity ***i***.

S2. This rod is placed in a region of space where there is a constant uniform magnetic field ***B***, also horizontal and which presents and angle ***θ*** with it.



S3. Knowing that ***i*** = 0.01 A, ***B*** = 0.3 T, ***l*** = 0.5 m, ***g*** = 9.8 m/s^2 and ***m*** = 0.045 Kg, what is the value of ***θ*** necessary for the rod to be in equilibrium?

Figure I. The task given to students

Results

Part One: Analysis of mistakes and their potential usefulness.

Mistakes made by students are reported and analyzed in this section. For this purpose, students' productions during the problem solving task are interpreted in terms of the activation of conceptual and frames. Also, other contexts are proposed in which the activation of the same resources could lead to correct answers.

Mistake type 1: inappropriate mapping of a conceptual resource

As an example, during the solving of problem 2, a conceptual resource was identified which was given the name *container*. This resource, useful in situations in which objects have to fit into containers, has been activated by most students and thus they interpret that the image of the person is contained in the mirror, and therefore the mirror has to be as large as the image to be seen. In the same problem, another resource is activated, which has been named *the farther, the smaller*. Activation of this resource leads students to state that as a person backs away from a mirror, the image is farther away and therefore it looks smaller.

Activation of *the farther, the smaller* together with *container* can account for students' verbalizations in which they state that as they back away from the mirror, a smaller mirror is needed. As an example, student "M" activates these two resources after S3 in problem 2:

"M": *… the mirror, to see himself completely, it should start at the floor, and be at least as tall as the person…*

"I" (interviewer): *is that what happens when you want to see yourself completely?*

"M": *hhmm, no, no…*

"I": *could you see yourself completely on a smaller mirror?*

"M": *well, that depends, where you're standing, I mean, the distance from the mirror… if you move forward, near the mirror, and the mirror doesn't reach the floor, you can't see your feet… but if you back away from the mirror, maybe a smaller mirror will be enough…*

Where does this incorrect answer come from? According to the approach described above, this could arise from an inappropriate mapping of the resource *the farther, the smaller* on the mirror situation. In other words, the apparent size of the image (which in fact is smaller when it is farther away) is compared to the actual size of the mirror as if its apparent size did not also change (the mirror is also farther away from a person backing away from it). This conjecture is depicted in Figure IIa. However, mapping this resource onto the image and the mirror simultaneously, can lead to a correct answer. The question that follows this observation is: is there a context in which this resource is spontaneously activated and mapped in such a way that it leads students to give a "correct" physical description? Figure IIb depicts one such situation, in which an observer views the exterior through a window. It seems reasonable to assume that these students have enough everyday experience to decide what can be seen through a window. Therefore, this same resource, but mapped differently, can be useful to reason about seeing objects through a window, as well as to reason about observing one's own reflection on a plane mirror.

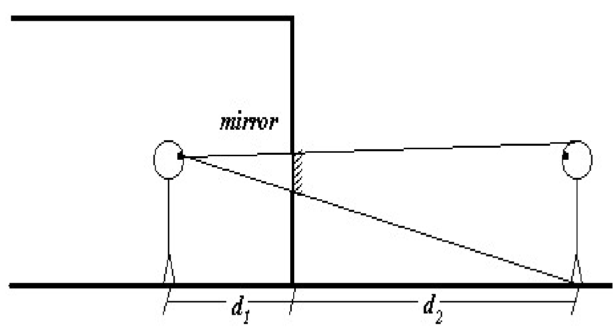


Figure IIa: "The farther, the smaller" inadequately mapped onto the mirror situation.

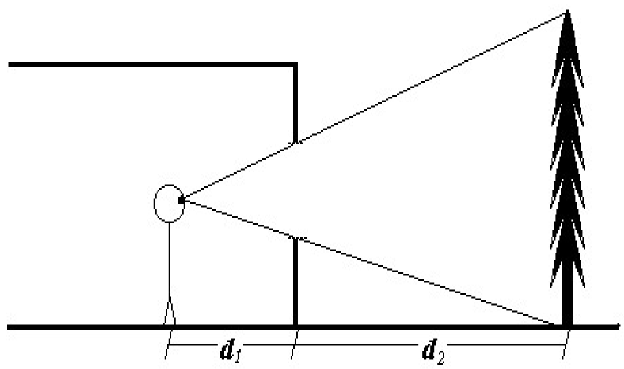


Figure IIb: "The farther, the smaller" mapped onto the window situation.

Mistake type 2: not productive activation of a conceptual resource

Another conceptual resource observed is the one named *eye contact*. This resource is useful to reason about two people seeing each other by means of a mirror, and to decide if they are making visual contact (each person can decide if the other one can see his eyes). "E" seems to have activated this resource, and mapped it onto himself and his image to decide about the smallest possible mirror

"E" (after S4):	*... they give me the person's height and how high his eyes are, so, it would have to be, at least, to see all of him, this high, that is as high as his eyes are, if…*
"I":	*so?*
"E":	*... it would have to be this high, the mirror, I mean, at least as high as his eyes, starting on his feet.*
"I":	*would you like to make any kind of drawing?*
"E":	*...no...*

Figure IIIa represents the activation of this resource to decide the size of the smallest mirror needed to see one's whole body image. Figure IIIb depicts the same resource in the context of deciding if a person is able to see another through a mirror.

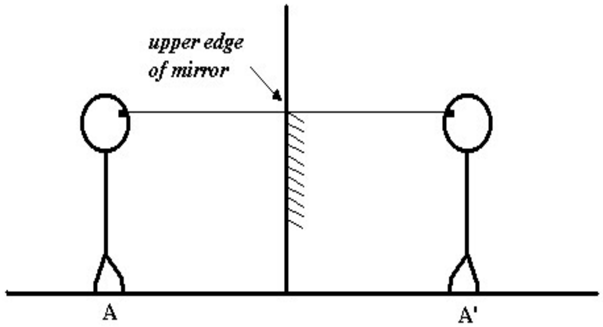


Figure IIIa: eye contact mapped onto a person and his/her image.

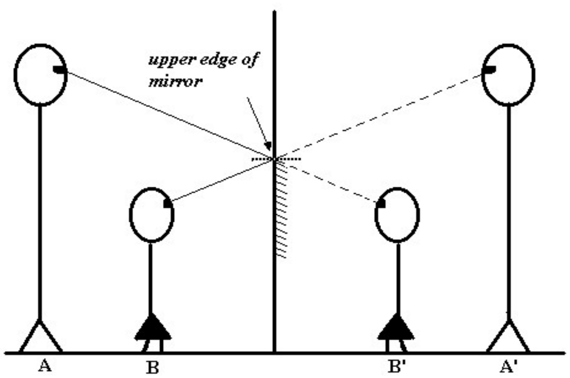


Figure IIIb: eye contact mapped onto two people to see each other.

Unlike the resource *the farther, the smaller*, the resource *eye contact* is not productive for deciding the size of the smallest mirror possible. This kind of mistake does not arise from mapping a useful resource inappropriately, as in the previous example, but rather from the activation of a resource that is not fruitful for the particular situation. Nevertheless, from an instructional point of view, it is fruitful (and will be discussed in the following section) to provide a context in which this resource is useful, such as the one depicted in Figure IIIb. This figure shows the minimum height a mirror should have for two people of different heights make eye contact through a mirror.

Problems 3 and 4 seem to induce the activation of another resource named *alignment*. Mapped on (electric and magnetic) dipolar moments, and (electric and magnetic) fields, respectively, this resource can lead students to provide physically accurate descriptions. This resource is seen to be activated in problem 3, onto the magnetic moment of the current circulating in the coil and the external magnetic field, leading students to give correct answers. However, in problem 4, this resource is mapped onto the conducting rod carrying a current and the magnetic field, which results in a physically incorrect description. Student "J" correctly solved problem 3, stating that the coil would not rotate due to the current induced by the variation of the field, because its dipolar moment was already oriented with the field. However, while solving problem 4, activates the same resource to provide an incorrect answer:

"I":	*(after the drawing in problem 4) what are you thinking about this?*
"J":	*that this external field will generate a torque... that makes the magnetic dipole on the conducting rod be aligned with it… and … that's it.*
"I":	*If you had to make up a question to this problem, what would it be?*
"J":	*Hmm.... to calculate the magnetic moment, of the rod... oh! I don't have the radius!*
"I":	*is the rod circular?*
"J":	*no, it isn't...*
"I":	*is there a magnetic moment?*
"J":	*no*
"I":	*then... is there anything going on? I mean, does that field have any effect on the rod?*
"J":	*no... except for aligning it… right?*
"I":	*you're saying it will align the rod, with what?*
"J":	*yeah... well, no, it doesn't do anything...*
"I":	*nothing?*
"J":	*well, no! 'cause...*
"I":	*So, you think nothing happens?*
"J":	*no! I think something does happen, but, hmmm, no... now I'm really not sure that anything actually happens…*
"I":	*but what do you think does happen?*
"J":	*Well, the thing about aligning it with the field, but that's for coils... I don't know, I'm confused now*

……………………

"J" (after S3)	*... in equilibrium!! But the field doesn't do anything to it!?...It is lying somewhere... but this rod isn't lying on any surface, is it? ¿WHERE is it?!?! I mean... is it in the air?!?!?!*
"I":	*yes*
"J":	*... honestly, I have no idea...*

"I": *forget about B for a second... what happens to a rod just placed in mid-air?*

"J": *it falls…and for what value will it be in equilibrium... no, I really don't know how to calculate this…. Can it be solved?*

As in the previous examples, this student's mistake comes from the activation of a resource that is not fruitful to address the situation. "J" activates *alignment* and not *balance* which, mapped on the forces acting on the conducting rod, could lead to a physically correct description.

The following is an excerpt from student *"C"* when she is solving problem 4):

"C" (after S3): *...I can compute the moment... the moment…*

"I": *which moment? Do you mean torque?*

"C": *no, torque is the product of the field times a "moment"...*

"I": *the dipolar magnetic moment?*

"C": *that's it! The dipolar magnetic moment...*

"I": *Is there a dipolar magnetic moment there?*

"C": *well, if there's a current, I guess there would have to be, right? Cause' if there weren't, there couldn't be a torque that makes the rod rotate (meaning a rotation towards the direction of the field)*

"I": *ok... but can we forget about the torque for just a second, and go back to the question of whether there is a magnetic moment?*

"C": *the moment was the product of the area times the current?*

"I": *yes, for a closed coil.*

"C": *...for a closed coil... oh!!! Right! It was for a closed coil... for a conductor... no wonder I couldn't come up with it using my right hand! … well, no! there is no moment...*

"I": *is there anything else there?*

"C": *...there has to be a force*

"I": *how so?*

"C": *a force tending to align the rod with the field.*

"I": *and what causes that force?*

"C": *obviously it has to be the external field... the current and the length of the rod…*

"I": *where is that force applied and what characteristics does it have?*

"C": *Well, the force is a vector, and B is also a vector, so it would have to be i times l times the sine of the angle, right?*

"I": *and where is that force pointing?*

"C": *hmmm, well, it would have to be perpendicular to the horizontal plane… upwards, the hmm, the force points up...*

"I": *so, how will that affect the rods movement?*

"C": *applying the force... it would have to lean towards the direction of the field...*

"I": *how?*

"C": *This is the rod… I can't do it in the other way with the right hand... the rod would have to go that way, right? … it would have to move to the left?*

"I": *how? Are we looking at the direction where the force is pointing?*

"C": *yeah, where the force is pointing... ok, well, I know that the force has to be perpendicular pointing up*

"I": *ok, then?*

"C": *but no, it's going to move it towards the direction of the field, and the value of theta... well, it's gonna have to be the sine of the angle with the intensity of the field.*

This protocol shows a mistake similar to the previous one. However, it exhibits an activation of the resource *alignment* which is more stable; since this activation persists even after "C" expresses that the force on the rod is perpendicular to the horizontal plane and points up. This mistake comes from the activation of a resource that, though unfruitful in this context, is useful in others.

Mistake type 3: not productive activation of a frame

In what follows a third kind of mistake is analyzed, related to the activation of frames. The excerpt presented corresponds to the protocol of "F", while solving problem

2. At first, this student activates the *qualitative sense-making frame*, and afterward activates the *quantitative sense-making frame* (Tuminaro, 2004):

"F":	*...ok, if he wants to see all his body, it would have to be... hum, I mean, ... it would have to be a large mirror, 1.65m or more... at least 1.65 to see himself completely, I mean, it also depends on the distance he is standing from the mirror... if he is too close, even if the mirror is very large, he will see "less"... and... well, maybe he can come up close to the mirror and look in some way so that he can see his feet… if, well, I mean, I'd place it a bit over his head...*
"I":	*Do you think this problem could be solved more... concretely?*
"F":	*What do mean "solve"? Make computations? Compute the height? Well, not as it is, I don't have the distance from the person to the mirror*
"I":	*And if you did have that distance?*
"F":	*...well, if I have the mirror here, and he is here, well, there I could compute that somehow... looking at the light rays more or less..*
"I":	*how?*
"F":	*If this is the mirror, and I take out rays from his head here to the end of the mirror, and the other ones to the other end... it's like... I think... if he's standing here like the drawing shows... and the mirror... there, (laughs softly) from the eye, I would have to cover his image completely, I mean the reflection, and that way I could come up with the size of the mirror, I think...*

The first part of F's answer is incorrect and, if the second part were not present, this could be viewed as a mistake due to inadequate mapping of the conceptual resources of *container* and *the farther, the smaller*. However, analyzing the complete protocol allows to understand that the mistake is also related to the activation of the *qualitative sense-making frame*, on the basis of which the answer to the problem does not involve algebraic or graphic computations, and only an argument based on previous (probably everyday) knowledge of mirrors and images. The interviewer's question regarding a more "concrete" solution seems to induce the activation of the *quantitative sense-making frame*, and thus the solution involves computations and/or graphic considerations. Nevertheless, the resource of qualitative solving is very useful to make qualitative predictions that can be later confirmed and compared to formal computations. Moreover, it is desirable and often absent in student's problem solving behavior, and its activation should not be disregarded even when it could lead momentarily to "incorrect" answers. Once again, both frames are useful in different contexts, and their activation can therefore lead to correct as well as incorrect answers. This example of F's protocol has been chosen to show how a mistake can come from the unproductive activation of a frame that in other contexts can be very useful.

Part Two: the knowledge of mistakes and their relation to instruction

The analysis of mistakes on the basis of cognitive resources can provide suggestive approximations to the problem of instruction. Two questions arise from this analysis: "what is the use of knowing what kind of mistake a student is making when solving a problem?" and "what is the use of knowing in what context the activation of the same resources could lead to correct answers?"

Understanding where students' mistakes come from allows to better knowing what it is that students do know and to be better prepared to work on that basis. For example, a mistake arising from the inadequate mapping of a conceptual resource that is potentially useful for the situation could require a different action than a mistake due to the activation of an unproductive conceptual or frame. An instructor's intervention should be different in each case because the "useful information". If the mistake comes from an inappropriate mapping, as in the case of the resources *container* and *the farther, the smaller* for problem 2, the comparison of this situation with another one such as a person looking out of a window (Figure IIb) could likely induce the activation of the same resources, and therefore the subject could compare the answer given in each situation. Since looking out of a window is a part of everyday experience for (almost) everyone, it is likely that the activation of these resources that naturally takes place in this context could result in an aid to address the mirror problem (it has been studied in more detail in Buteler & Coleoni, 2009).

If the mistake observed comes from the activation of a resource which is unproductive to solve the problem, as in the case of "eye contact" or "alignment", an efficient strategy could be to have induced the comparison with the answers given in contexts where such resources are productive, analyzing similarities and differences. Such a comparison could lead students to "learn" in what contexts those resources are productive and why the characteristics of other contexts make this resource unproductive.

In any case, the comparison is made between the reasoning of one same subject in different contexts, and not between the student's and the teacher's or a textbook. These strategies do not foster a barrier between the student's thinking and the "correct" reasoning, because they allow reinforcing students' ideas in the appropriate contexts. These comparison strategies, however, require certain knowledge of contexts in which students could potentially activate fruitful resources. Teachers' expertise in physics instruction together with studies aimed at testing the effectiveness of such strategies could be of great value.

Finally, mistakes coming from an unproductive activation of a frame could call for more extended actions sustained through time. Comparison strategies between “close” contexts as the ones presented in this paper could be insufficient for this purpose. Close contexts are those sharing the kind of task, the social or physical environment, and that differ only in the physical situation presented. Far contexts are those in which the tasks presented differ more radically from each other (such as problem solving vs. argumentation for or against a thesis) or in which the social environment is different (classroom vs. informal interviews), etc. It is likely that the changes in context needed to foster the activation of productive or unproductive frames be more pronounced than those needed for conceptual resources. The authors intend to address these issues in the future.

Discussion

The theoretical framework adopted has allowed us to interpret students' verbalizations. This analysis allows understanding possible mechanisms by which students produce physically “incorrect” answers. The interpretation of protocols in terms of conceptual resources activated and mapped in different contexts can account for two mistakes of a different nature. One of them (mistake type 1) occurs when the resource activated is useful to address the situation, but has been mapped in such a way that leads to contradict a physically correct result. Such is the case of *the farther, the smaller* in problem 2. The other kind of mistake takes place when the resource activated is not fruitful to address the situation (mistake type 2). Such is the case of *alignment* in problem 4, or of *eye contact* in problem 2. This analysis of mistakes enables to imagine contexts in which a productive activation of these same resources could occur. Comparison between these two situations could lead students to learn in a way that is tuned with what they already know. They are not lead to disregard their knowledge when it is incorrect, but rather to refine the way in which they reason with the cognitive tools they do have. These findings also bring up the question of how exposing students to contexts in which the resources activated are useful could foster metacognitive processes that enable them to reorganize what they already know. These questions are being approached by the authors at present.

Regarding frames, a suggestive result is the variability in their activation by one same student. Such is the case of student “F” in problem 2 (mistake type 3). This indicates that students have epistemic capacities potentially useful to address physics problem solving, and opens the question of how to make the best use of these abilities, instruction-wise.

In general terms, the present study aims at showing how a more detailed analysis of students' mistakes can change the view of teaching and of research. The view of teaching changes because a new meaning is assigned to students' “incorrect” answers, and they can be regarded as valuable. As for research, new questions arise that lead to hypothesis regarding the efficiency of comparison strategies aimed at fostering students' learning on the basis of what they know.

References

Becerra Labra, C., Gras-Martí, A. & Martínez-Torregrosa, J. (2004). Análisis de la resolución de problemas de física en secundaria y primer curso universitario en Chile. *Enseñanza de las Ciencias, 22*(2), 275-286.

Bring, T. & Redish, E. (in press). Analyzing Problem Solving Using Math in Physics: Epistemic Framing via Warrants. *American Journal of Physics- PER Secction*.

Buteler, L. & Coleoni, E. (2009). Cómo aprovechar la naturaleza contextual del conocimiento para resolver un problema de física: un abordaje basado en recursos cognitivos. *Investigaçoes em Ensino de Ciencias*, 14(1), 7-24

Chi, M. (2005). Commonsense conceptions of emergent processes: why some misconceptions are robust. *Journal of the Learning Sciences, 14*(2), 161-199.

Chi, M., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121- 152.

Chi, M., Glaser, R., & Rees, E. (1982). Expertise in Problem Solving. In Stenberg, R. (Ed.). *Advances in the Psychology of Human Intelligence* (pp 7-75). New Jersey: Lawrence Erlbaum Associates.

diSessa, A. & Sherin, B. (1998). What changes in conceptual change? *International Journal of Science Education, 20, 20* (10), 1155-1191.

diSessa, A. (1993). Toward an epistemology of physics. *Cognition and Instruction, 10*(2), 105-225.

Foster, T. (2000). The development of student problem-solving skills from instruction emphasizing qualitative problem solving. PhD Minnesota University USA.

Gil Pérez, D. (2003). Constructivism in science education: The need for a clear line of demarcation. In Psillos, D., Kareotoglo, P., Tselfes, V., Hatzikraniototios, E., Fassoulpoulos, G. & Kallery, M. (Eds). *Science Education Research in the Knowledge-Based Society* (pp 9-17). Netherlands: Kluwers Academic Publishers.

Gil Pérez, D. & Martínez Torregosa, J. (1983). A model for problem solving in accordance with scientific methodology. *European Journal of Science Education, 5, 5* (4), 447-455.

Guisasola, J., Furió, C., Ceberio, M. & Zubimendi, J. (2003). ¿Es necesaria la enseñanza de contenidos procedimentales en cursos introductorios de física en la Universidad? *Enseñanza de las Ciencias,* Número Extra, 17-28.

Hammer, D. (2004). The variability of student reasoning, lecture 3: manifold cognitive resources. In Redish, E. and & Vicentini, M. (Eds). *Proceedings of Enrico Fermi Summer School* (pp 321-340). Bologna: Società Italiana di Física.

Hammer, D., Elby, A., Scherr, R. & Redish, E. (2005). Resources, framing and transfer. In Mestre, J. (Ed). *Transfer of learning from a modern multidisciplinary perspective* (pp 89-119). Greenwich, CT: Information Age Publishing.

Heller, K. & Heller, P. (1995). *The competent problem solver,* Calculus version. Minneapolis, MN: Author.

Huffman, D. (1994). The effect of explicit problem solving instruction on students conceptual understanding of Newton's' Law. PhD Minnesota University USA.

Larkin, J., McDermott, L., Simon, D. & Simon, H. (1980). Expert and novice performance in solving physics problems. *Science, 208*, 1335-1342.

Leonard, W., Gerace, W. & Dufresne, R. (2002). Resolución de problemas basada en el análisis: hacer del análisis y del razonamiento el foco de la enseñanza de la física. *Enseñanza de las Ciencias, 20* (3), 387-400.

Maloney, D. (1994). Research on problem solving: physics. In Gabel, D. (Ed). *Handbook on Research of Science Teaching and Learning* (pp 327-354). New York: Mc Millan Publishing Company.

Mestre, J., Dufresne, R., Gerace, W., Hardiman, P. & Tonger, J. (1993). Promoting skilled problem-solving behavior among beginning physics students. *Journal of Research in Science Teaching, 30* (3), 303-317.

Pozo, J. (1996). No es oro todo lo que reluce ni se construye (igual) todo lo que se aprende: contra el reduccionismo constructivista. *Anuario de Psicología, 69*, 127-139.

Redish, E. (2004). A theorical framework for physics education research: modeling student thinking. In Redish, E. & Vicentini, M. (Eds). *Proceedings of Enrico Fermi Summer School* (pp 1-63). Bologna: Società Italiana di Física.

Russ, R., Scherr, R., Hammer, D. & Mikeska, J. (2008). Recognizing mechanistic reasoning in student scientific inquiry: A framework for discurse analysis developed from philosophy of science. *Science Education, 92,* 499-525.

Tuminaro, J. (2004). A cognitive framework for analyzing and describing introductory students' use and understanding of mathematics in physics. PhD University of Maryland USA.

Tuminaro, J. & Redish, E. (2007). Elements of a cognitive model of physics problem solving: Epistemic games. *Physical Review Special Topics – Physics Education Research*, *3*, 020101.

Factors That Influence Sense of Place as a Learning Outcome and Assessment Measure of Place-Based Geoscience Teaching

Steven Semken
Carol Butler Freeman
Nievita Bueno Watts
Arizona State University

Jennifer J. Neakrase
New Mexico State University

Rebecca Escobar Dial
Dale R. Baker
Arizona State University

Abstract

Sense of place encompasses the meanings that a given place holds for people and the attachments that people develop for that place. Place-based science teaching integrates the natural and cultural meanings of a place as context for scientific study, and hence leverages the senses of place of students and instructor. It has been proposed that this method enhances relevance and interest for introductory students, particularly those with cultural ties to the places under study. Authentic evidence of place-based learning comprises not only gains in locally situated knowledge and skills, but also enrichment of the sense of place. Valid and reliable surveys for measuring sense of place exist and have been tested successfully as assessment instruments. However, a student's proximity of residence and history of visitation with a place used as the setting for a lesson may also influence his or her sense of that place. To investigate the possible effects of these factors and further explore the sense of place in assessment, introductory geology students were surveyed on their proximity of residence to, history of visitation to, and sense of Grand Canyon: an iconic place and the subject of a class laboratory exercise. Frequency and recency of visits to Grand Canyon, but not proximity of residence to it, were correlated with student's sense of place. These findings suggest that place-based geoscience teaching is applicable to nonresident and local students alike, but that prior experiences with the place may influence a student's receptivity to the method.

Correspondence should be addressed to Steven Semken, Arizona State University, School of Earth and Space Exploration, POB 871404, Tempe, AZ 85287-1404, semken@asu.edu

Sense of Place in Science Teaching

Place is defined as any locality or space that has become imbued with meaning by human experience in it (Tuan, 1977). A spectrum of humanistic and scientific meanings may accrue to any given place, reflecting all of the ways that diverse individuals and groups know and experience it. People also tend to build strongly emotional attachments to meaningful places. The *sense of place* has been defined as the combined set of the *place meanings* and *place attachments* that a person or a group develop for a place

(Brandenburg & Carroll, 1995; Williams & Stewart, 1998). Sense of place therefore encompasses the cognitive and affective domains, and possibly also the psychomotor domain if particular kinesthetic activities are associated with or localized in a particular place (Semken & Butler Freeman, 2008). The nature of sense of place and its relevance to science education are discussed in detail in a paper by Semken and Butler Freeman (2008).

Place-based (sometimes called *place-conscious*) teaching (Woodhouse & Knapp, 2000; Smith, 2002; Gruenewald, 2003a, 2003b; Sobel, 2004; Gruenewald & Smith, 2008) is a situated approach that consciously leverages (Lim & Calabrese Barton, 2006) and enriches the senses of place of students and instructor through localized experiential learning, cross-cultural and trans-disciplinary content and pedagogy, and outreach to the community. In contrast, a great deal of conventional teaching is decontextualized and focused on a canonical list of abstract principles or isolated facts (Barab & Roth, 2006), of which only a few may have any local significance.

The Earth, ecological, and environmental sciences are taught in and by means of places. Place-based teaching in the natural sciences, offering meaningful context and practical relevance (Aikenhead, 1997, 2001; Semken & Morgan, 1997; Butler, Hall-Wallace, & Burgess, 2000; Semken, 2005; Glasson, Frykholm, Mhango, & Phiri, 2006; Chinn, 2006; Semken & Butler Freeman, 2008), is thought to improve engagement and retention of students, particularly for members of indigenous or historically rooted communities who already have rich senses of the places under study (Cajete, 2000; Emekauwa, 2004; Riggs, 2005; Aikenhead, Calabrese Barton, & Chinn, 2006; Levine, González, Cole, Fuhrman, & Le Floch, 2007). Conversely, teaching that contradicts or minimizes such students' senses of place may dissuade them from studying science (Kawagley, D. Norris-Tull, & R. A. Norris-Tull, 1998; Aikenhead & Jegede, 1999; Semken, 2005; Chinn, 2006). This may be particularly true for the geosciences, which by their nature penetrate and probe the physical substrates of places that are deeply meaningful or even sacred to some cultural groups. Geoscience educators should be aware and respectful of possible pre-existing place attachments among their students, particularly when teaching in the field or about certain topics, such as mining, recreation, and other forms of resource extraction or use (Semken, 2005).

To this point, research on the effectiveness of place-based teaching has been focused on elementary and secondary school programs and has yielded affirmative but indirect results, which include:

- significantly enhanced student performance on standardized multi-disciplinary achievement tests (Lieberman & Hoody, 1998);

- significantly improved student achievement motivation (Athman & Monroe, 2004) and critical-thinking skills (Ernst & Monroe, 2004); and

- more collaborative and interdisciplinary practice, and more frequent use of service-learning projects, by teachers (Powers, 2004).

Although each of these studies endorses place-based teaching, none directly addresses the defining attribute and aim of the approach, which is intimate, meaningful, and sustainable engagement with the surrounding natural and cultural environments (Lim & Calabrese Barton, 2006; Ault, 2008). Authentic evidence of place-based learning should thus encompass not only significant improvement in locally situated content knowledge and skills, but also significant enrichment or enhancement of the sense of place, which encapsulates the student's personal connection to the study place or places (Semken & Butler Freeman, 2008). Hence, authentic assessment of place-based teaching would be facilitated by any valid and reliable means of measuring sense of place in students.

Deconstructing and Measuring the Components of Sense of Place

Personal senses of a given place can vary greatly, and Relph (1976) has described how these can be ranked by their depth or intensity, from utter alienation ("existential outsideness," p. 51) to complete belonging ("existential insideness," p. 55). Hence it is possible to quantitatively measure an individual's sense of a particular place. Such measurement finds application in land-use planning and resource management, in which it is now often necessary to account for the senses of place of different stakeholders (Williams & Stewart, 1998; Clark & Stein, 2003). Quantitative analysis of the sense of place is also important to the recreational and tourism industries (Bricker & Kerstetter, 2002). As a consequence, the construct has been extensively characterized in environmental psychology, and there now exist published psychometric instruments designed to measure each of the two principal components of sense of place: place attachment and place meaning.

Place Attachment

Place attachment is an emotional bond to a place that develops from direct experience (e.g., living, working, or vacationing in the place), vicarious engagement (e.g., through books or visual media), or some combination thereof (Relph, 1976; Williams & Stewart, 1998). Love of one's hometown or a favorite campsite; a desire to protect a wilderness area or a historic urban structure from demolition; delight in collecting and viewing paintings made of a landscape or region one may or may not have ever visited: each is an example of place attachment.

Shamai (1991) proposed a seven-point empirical intensity scale for place attachment, based on Relph's (1976) ranking system, ranging from no sense of place at one extreme, to a willingness to make personal sacrifices on behalf of a place at the other. Shamai's test of this scale on students in a Jewish school in Toronto, and a separate use of the scale by Kaltenborn (1998) to characterize place attachment among inhabitants of the Svalbard archipelago, showed that an empirical instrument could resolve and measure intensities of place attachment in two geographically and culturally distinct groups.

A valid and more generalizable place-attachment survey was developed by Williams and colleagues (Williams, Patterson, Roggenbuck, & Watson, 1992; Williams & Vaske, 2003). In accord with a theoretical model from environmental psychology

(Brown, 1987; Williams et al., 1992), their instrument measures two dimensions of place attachment: *place dependence*, the capacity or potential of a place to support an individual's needs, goals, or activities (Stokols & Shumaker, 1981; Williams & Vaske, 2003); and *place identity*, an individual's various affective relationships to a place (Proshansky, 1978; Proshansky, Fabian, & Kaminoff, 1983; Korpela, 1989; Williams & Vaske, 2003), such as memories, preferences, and feelings. Williams and Vaske validated this instrument using data from 2819 respondents polled at six recreational sites and parklands in Colorado and Virginia, and on a university campus in Illinois. Their study, detailed in their 2003 paper, confirmed construct validity with a factor analysis that sustained the two-dimensional model of place attachment; and convergent validity as significant positive correlations between the two dimensions and theoretically linked variables, such as familiarity and frequency of visitation.

Williams and Vaske also showed that a concise survey with no more than six place-dependence items and six place-identity items (Table I) had good internal-consistency reliability (Cronbach's alphas ranging from 0.81 to 0.94) across all seven study places, and could be considered highly generalizable (coefficients 0.924 for place dependence and 0.869 for place identity) to different places. Additional items yielded little improvement in generalizability (Williams & Vaske, 2003). The survey uses a five-point Likert scale. In this paper it will be identified as the *Place Attachment Inventory* (*PAI*).

Table I
Place Attachment Instrument of Williams & Vaske (2003)

I feel (place name) is a part of me.
(Place name) is the best place for what I like to do.
(Place name) is very special to me.
No other place can compare to (place name).
I identify strongly with (place name).
I get more satisfaction out of visiting (place name) than any other.
I am very attached to (place name).
Doing what I do at (place name) is more important to me than doing it in any other place.
Visiting (place name) says a lot about who I am.
I wouldn't substitute any other area for doing the types of things I do at (place name).
(Place name) means a lot to me.

The things I do at (place name) I would enjoy doing just as much at a similar site.

Note. The odd-numbered items measure place identity, the even-numbered items measure place dependence, and the final item is reverse scored. This instrument is used with a Likert scale in which 1 corresponds to "strongly agree," 2 to "agree," 3 to "neutral," 4 to "disagree," and 5 to "strongly disagree."

Place Meaning

Although the meanings that imbue places run the gamut from spiritual (e.g., sacredness) to scientific (e.g., interpretation of bedrock geology), place meaning is always contextually bound to the place itself. Therefore, to be authentic, any psychometric measure of place meaning should be developed empirically and locally, with items emergent from the set of meanings held by those who variously inhabit, promote, visit, or consider the place. The work of Young (1999), who created an empirical place-meaning survey for a World Heritage parkland in northeast Queensland, Australia, exemplifies the construction and valid use of this kind of instrument. A tourism geographer, Young described place meanings as socially constructed and negotiated between those who "produce" and disseminate them, such as tour guides and interpretative specialists; and those who "consume" (hold or construct) them, such as tourists and other visitors. This model is relevant to place-based formal education, in that teachers can be described as "producers" and students "consumers" (although one would expect more of a two-way exchange of place meanings in this more open and collaborative learning environment). Young's model for construction of place meanings is also analogous to those of other theorists of sense of place (Ryden, 1993; Casey, 1996).

Young extracted a set of produced meanings from a textual analysis of brochures published to promote the region, and surveyed tour operators to determine which of these were most important. A set of consumed meanings emerged from brief semi-structured interviews of visitors in the parks. Young then incorporated these parallel sets of place meanings into a 30-item questionnaire (Table 2) with a five-point scale, which polls respondents on whether each of the items is a poor, fair, good, very good, or excellent description of the place. Young used this instrument in a study of different influences on the place meanings held by tourists. One finding particularly relevant to place-based teaching was that respondent place meanings were influenced by the level of prior knowledge of the place, preferences for particular types of surroundings, and sociocultural background.

Table II
Place Meaning Instrument of Young (1999)

Ancient	Privilege to visit	Fun
Pristine	Relaxing	Threatened

Scenic	Important for Aboriginal culture	Crowded
Beautiful	Overdeveloped	Dangerous
Remote	Tropical	Interesting
Unique	Unusual	Educational
Important to preserve	Scientifically valuable	Tranquil
Authentic	Ecologically important	Spiritually valuable
Fragile	Wilderness	Historical
Exotic	Adventurous	Comfortable

Young (1999) did not report on the validity or reliability of this instrument. In the absence of statistical data, this survey can be considered valid for the measurement of local place meaning in individual respondents on the basis of Young's theoretically sound model for the construction of place meanings and the naturalistic, empirical method by which the survey was created, following lines of reasoning put forth by Mishler (1990), Aikenhead and Ryan (1992), and Semken and Butler Freeman (2008). Because the survey was developed for use in Australia, Young also did not address its generalizability to other countries. However, nearly all of the items are generic enough to be applicable to other parklands or wild places elsewhere. This instrument will be referred to in this paper as *Young's Place Meaning Survey* or *YPMS*.

Applications to Assessment of Place-Based Science Teaching

In a recent preliminary study, Semken and Butler Freeman (2008) used the PAI and YPMS as pre- and post-tests of sense of place in a group of 27 students who completed an experimental Arizona-based, culturally inclusive, meaning-rich, introductory geology course at a large state university in metropolitan Phoenix. They ran dependent-samples *t* tests on the pre-test and post-test means for the PAI and YPMS, and observed significant ($p < 0.01$) increases in mean student place attachment and place meaning for Arizona between the start and completion of the place-based course. Semken and Butler Freeman's results suggest that the PAI and YPMS are generalizable and sensitive enough for use as assessment tools. However, since a control group was not available for this study, the effectiveness of the place-based course in enhancing sense of place was not conclusively shown. Neither could this study address any subjective factors (e.g., familiarity with or prior experiences in the study place) that might be predictors of individual differences in sense of place. Such factors would likely influence the effectiveness of this teaching approach in any large academically, ethnically, culturally, socioeconomically, and geographically diverse student population, such as the typical large-enrollment ($n > 100$) introductory geoscience classes that universities regularly offer.

The study described in this paper is an exploration of several factors that are likely to be present to some degree in all introductory geoscience students, and which may be correlated with place attachment or place meaning. These factors may influence the use of sense of place (more specifically, sense of the specific place or places examined in the curriculum) as a learning outcome or a metric of the effectiveness of a place-based approach to geoscience teaching.

Factors Thought to Influence a Student's Sense of Place

Proximity to a Place

In any large university geoscience course, some of the students will be local and others will hail from outside of the region. How will these different groups respond to teaching that is explicitly situated locally? One can certainly develop a rich sense of a place without ever coming near to it (Relph, 1976; Proshansky, Fabian, & Kaminoff, 1983). Consider the Western novels of the author Karl May (1842—1912), who never ventured west of New York state (Wohlgschaft, 1994), but who proffered meanings and instilled strong attachments to the western USA and its indigenous cultures in several generations of his fellow Germans. However, in general, it would be expected that place attachment and place meaning would be associated with familiarity derived from the proximity of a student's residence to the place. Familiarity could arise from residing in or near the place, from regularly traveling through or nearby to it, or from hearing or seeing the place referenced frequently in local media, schools, museums, or even casual conversation. A sense of place thus constructed could either be affirmative (e.g., feelings of community) or negative (e.g., boredom with the place) (Pretty, Chipuer, & Bramston, 2003). Young (1999) found that a respondent's place of origin was the factor most strongly correlated with place meaning in his Australian study: domestic visitors to the study region scored higher on the YPMS than did visitors from overseas. His interpretation was that the former were more familiar with the area owing to well-publicized environmental disputes about a decade earlier (Young, 1999).

Visits to a Place

Individuals who reside far from a place may still make frequent visits to it for avocational or vocational reasons. Frequent visitation, motivated by place dependence, may in turn bolster the visitor's place identity (Moore & Graefe, 1994), and thus enhance place attachment. In a study conducted in four wilderness areas, Williams, Patterson, Roggenbuck, & Watson (1992) found attachment to these places to be strongly associated ($p < 0.001$) with the number of a respondent's previous visits, and also with the number of years since the first visit (i.e. the length of the history of visitation). The effect of prior visitation on place meaning is less clear. Young (1999) found that frequency of visits to natural environments in general is associated ($p < 0.01$) with richness of place meaning for the tourist respondents in his study region, but not with repeat visits to the place itself. This unexpected result could have been a consequence of the temporal and spatial constraints on free exploration imposed by guided tours, which were used by the majority of the respondents (Young, 1999).

The Study

Research Question Addressed

If pre-post or formative changes in a student's sense of place are to be used as an assessment measure in place-based science education (Semken & Butler Freeman, 2008), any pre-intervention meanings or attachments the student has for the place(s) under study must be understood and accounted for. Hence, the research question addressed in this study: Is a student's level of prior experience (measured as the proximity of residence and history of visitation) with a place that serves as the subject of a place-based geoscience intervention correlated with the student's prior sense of that place (measured as intensity of place attachment and richness of place meaning)?

Setting

The study was carried out in an introductory physical geology laboratory course during the spring 2005 semester. Most students in this course are not science majors, and they commonly enroll to fulfill a general studies requirement for graduation. The typical spring enrollment for this course is approximately 1100 students, who register in lab sections of no more than 30 students each to fit their class schedules. Each lab section meets for a two-hour session each week for 14 weeks (12 laboratory-room sessions, one on-campus field trip, and one research session held in the university map library). The course is inquiry–driven, systematic, and well-organized; each week's activities are outlined in detail in a custom-published laboratory manual (Reynolds, Johnson, & Stump, 2005) that each student purchases in advance. The content of the course emphasizes the physical landscapes of Arizona and the geology that underlies them.

The study centered on the ninth-week laboratory class in this course, which is focused on the geology of Grand Canyon in northern Arizona. Other places in Arizona are addressed in other weeks and other chapters of the manual, but Grand Canyon was selected because of its exceptionally rich place meanings, its importance to many diverse groups throughout history (Hirst, 2006; Powell, 1895/1987; Pyne, 1998; Morehouse, 1996; Beus & Morales, 2003) and its general recognizability, even to those who have never been there. The objective was to maximize any potential prior effects on student senses of place by selecting the most iconic place used in the Arizona-based curriculum.

Population

Race, ethnicity, and sex were not tabulated within the study population, but it appeared to be reasonably representative of the undergraduate student population at the university during spring 2005: 53% female, 47% male; 69.2% White, 5.1% Asian-American, 3.7% African-American, 2.2% Native American; 12.9% Hispanic; 2.7% international; and 4.2% undeclared or unknown. Approximately 400 students participated in the survey.

Survey

The first part of the survey used in this study consisted of four multiple-choice items used to determine a student's proximity of residence to and history of visitation of Grand Canyon (Table III). Proximity of residence was expressed as approximate driving time from the respondent's home to Grand Canyon. It was thought that respondents, if aware of their proximity to Grand Canyon, would more accurately know the driving time than the actual distance in miles or kilometers.

Table III
Survey Items Relating to Proximity to and Visitation of Grand Canyon

Of all the places you have lived for at least one year, what was the shortest amount of driving time between your home and the Grand Canyon? (Possible responses: less than 3 hours; 3—6 hours; or more than 6 hours)
How many times have you visited the Grand Canyon in total? (Possible responses: zero; 1—3 times; or more than 3 times)
How many times have you visited the Grand Canyon in the last year (12 months)? (Possible responses: zero; 1—3 times; or more than 3 times)
How long ago was your last visit to the Grand Canyon? (Possible responses: never; within the last year; or more than 1 year ago)

The responses to the questions shown in Table III were selected after considerable debate by the authors. These ranges were defined in order to reflect geographic and personal factors, and to allow for enough categories to elicit a variety of responses from the students. The range for driving time, with intervals ending and starting at three and six hours, reflects the roughly three-hour driving time from the university to Grand Canyon and the roughly six-hour driving time from the farthest places in our state to Grand Canyon. The range for number of total visits and visits within the last year was intended to differentiate among students who had never visited Grand Canyon, who had visited only a few times, and who were frequent visitors. Similarly, the range for length of time since the last visit to Grand Canyon was intended to distinguish those who had never visited the place, those who visited it some time ago, and those who visited it recently.

The second part of the survey consisted of the twelve PAI items as they are presented in Table I, verbatim from the published instrument of Williams and Vaske (2003), with "Grand Canyon" inserted as the place name. Students were asked to rate each statement on a five-point Likert scale, with 1 corresponding to "strongly agree," 2 to "agree," 3 to "neutral," 4 to "disagree," and 5 to "strongly disagree." For the first eleven items, a lower rating indicates a stronger place attachment; for the final item the opposite is true, so this item was reverse scored. A PAI score is calculated as the total of all twelve responses. Therefore the lowest PAI score, representing strongest place attachment, is 12; a neutral score is 36; and the highest score, representing weakest place attachment or place aversion, is 60.

The third part of the survey consisted of 27 YPMS items from the survey of Young (1999; Table II). Three place meanings from the original instrument (“tropical,” “fun,” and “comfortable”) were omitted, and the term “Aboriginal” was changed to “Native American,” to render the survey more locally relevant. Students were asked to rate the degree to which each of the 27 place-meaning terms represented Grand Canyon for them, on a five-point Likert scale identical to that used with the PAI. Strong agreement (expressed by a numerically low rating) with any of the terms except four (overdeveloped, threatened, crowded, and dangerous) indicates that Grand Canyon strongly holds that particular affirmative place meaning for the student. In the case of the other four terms, the opposite was held to be true, as these are meanings indicative of degradation of Grand Canyon. The YPMS score is calculated by summing the numerical responses to all items, with the four negative items reverse scored. The lowest YPMS score of 27 indicates that Grand Canyon holds the richest meanings for a student, whereas a score approaching the maximum of 135 indicates that Grand Canyon has little meaning to the student.

The survey was administered to the students in class one week before the scheduled Grand Canyon laboratory exercise. Participation in the surveys was voluntary, and the surveys were coded to maintain the anonymity of the participants. The students needed about ten to fifteen minutes to complete the surveys.

Data Analysis

Proximity and visitation versus place meaning and place attachment

In the analyses discussed below, proximity and visitation factors, indicated by responses to the four multiple-choice items at the head of the survey (Table III), were the independent variables; and PAI (Table I) and YPMS (Table II) scores were the dependent variables.

Place attachment (PAI score) versus proximity of residence to Grand Canyon

A one-way analysis of variance (ANOVA) was conducted to evaluate the hypothesis that student’s place attachment to Grand Canyon would be more strongly affirmative, on average, the closer that student lives to Grand Canyon. The independent variable, the proximity factor, comprised the three levels described above: less than 3 hours driving time, 3—6 hours driving time, and more than 6 hours driving time. The dependent variable was the individual’s PAI score. The ANOVA was non-significant, $F(2, 375) = 1.66$, $p = 0.19$. Table IV shows the means and standard deviations for PAI score for each level of the factor.

Table IV
Means and Standard Deviations of Place Attachment (PAI) Score for the Proximity Factor

Proximity Group	*N*	*M*	*SD*
Less than 3 hours	92	46.82	9.46
3—6 hours	224	45.02	8.82
More than 6 hours	62	46.63	9.30

Place attachment (PAI score) versus total number of visits to Grand Canyon

A one-way ANOVA was conducted to evaluate the hypothesis that student's place attachment to Grand Canyon would be more strongly affirmative, on average, the more times that student has visited Grand Canyon. The independent variable, the total visit frequency factor, included the three levels discussed above: never visited, visited 1—3 times, and visited more than 3 times. The dependent variable was the student's PAI score. The ANOVA was significant, $F(2, 383) = 23.70$, $p < 0.05$. The strength of the relationship between the total number of times visiting Grand Canyon and PAI score, as assessed by η^2, was small, with the total visit frequency factor accounting for 11% of the variance of the dependent variable.

Follow-up tests were conducted to evaluate pairwise differences among the means. Because the variances among the three groups ranged from 8.42 to 9.99 it was assumed that the variances were homogeneous, and post-hoc comparisons were made using the Least Significant Difference (LSD) test, which is appropriate for three levels of a factor. There were significant differences in the means between all of the groups (all $p < 0.05$). The group that had never visited Grand Canyon showed a weaker PAI score in comparison to the group that visited one to three times in total, and in comparison to the group that visited more than three times in total. The group that visited a total of one to three times in total also showed a weaker PAI score in comparison to the group that visited more than three times in total. The 95% confidence intervals for the pairwise differences, as well as the means and standard deviations for each level of the factor, are shown in Table V.

Table V
Means, Standard Deviations, and 95% Confidence Intervals of Pairwise Differences for Place Attachment (PAI) Score for the Total Visit Frequency Factor

Total visit frequency group	*N*	*M*	*SD*	Zero times	1—3 times
Never visited	148	48.66	8.43		
1—3 times	197	45.11	8.42	1.70 to 5.39*	
More than 3 times	41	38.46	9.99	7.21 to 13.18*	3.74 to 9.55*

Note. An asterisk indicates that the 95% confidence interval does not contain zero, and therefore the difference in means is significant at the 0.05 level using the LSD procedure.

Place attachment (PAI score) versus number of visits to Grand Canyon within the last year

A one-way ANOVA was conducted to evaluate the hypothesis that student's place attachment to Grand Canyon would be more strongly affirmative, on average, the more times that student visited Grand Canyon within the last year. The independent variable, the one-year frequency factor, included the three levels discussed above: never visited within the last year, visited 1—3 times within the last year, and visited more than 3 times within the last year. The dependent variable was the student's PAI score. The ANOVA was significant, $F(2, 378) = 11.57$, $p < 0.05$. The strength of the relationship between the number of times visiting Grand Canyon within the last year and PAI score, as assessed by η^2, was small, with the visit frequency factor accounting for only 5.8% of the variance of the dependent variable.

As above, follow-up tests were conducted to evaluate pairwise differences among the means. Because the variances among the three groups ranged from 8.83 to 11.31 it was assumed that the variances were homogeneous, and post-hoc comparisons were made using the LSD test. There was a significant difference in the means between not visiting Grand Canyon in the last year and visiting one to three times in the last year ($p < 0.05$). There was also a significant difference between not visiting in the last year and visiting more than three times in the last year ($p < 0.01$). No significant differences were seen between visiting one to three times in the last year and visiting more than three times in the last year ($p = 0.07$). The group that had not visited in the last year showed weaker place attachment in comparison to the group that visited one to three times and in comparison to the group that visited more than three times. The 95% confidence intervals for the pairwise differences, as well as the means and standard deviations for each level of the factor, are shown in Table VI.

Table VI
Means, Standard Deviations, and 95% Confidence Intervals of Pairwise Differences for Place Attachment (PAI) Score for the One-Year Frequency Factor

One-year frequency group	*N*	*M*	*SD*	Zero times	1—3 times
Zero times	339	46.45	8.83		
1—3 times	40	40.53	9.09	3.01 to 8.84*	
More than 3 times	2	29.00	11.31	5.09 to 29.81*	-1.11 to 24.16

Note. An asterisk indicates that the 95% confidence interval does not contain zero, and therefore the difference in means is significant at the 0.05 level using the LSD procedure.

Place attachment (PAI score) versus length of time since last visit to Grand Canyon

A one-way ANOVA was conducted to evaluate the hypothesis that student's place attachment to Grand Canyon would be more strongly affirmative, on average, the more recently that student has visited Grand Canyon. The independent variable, the recency factor, included the three levels discussed above: never visited, visited within the last year, and visited more than one year ago. The dependent variable was the student's PAI score. The ANOVA was significant, $F(2, 379) = 17.50$, $p < 0.05$. The strength of the relationship between how recently someone has visited Grand Canyon and PAI score, as assessed by η^2, was small, with the recency factor accounting for only 8.5% of the variance of the dependent variable.

Follow-up tests were conducted to evaluate pairwise differences among the means. Because the variances among the three groups ranged from 8.46 to 9.41 it was assumed that the variances were homogeneous, and post-hoc comparisons were made using the LSD test. There were significant differences in the means between all groups of length of time since visiting Grand Canyon (all $p < 0.05$). The group that has never visited showed a weaker place attachment in comparison to the group that visited within the last year and in comparison to the group that visited more than one year ago. The group that visited within the last year showed a stronger place attachment in comparison to the group that had visited more than one year ago. The 95% confidence intervals for the pairwise differences, as well as the means and standard deviations for each level of the factor, are shown in Table VII.

Table VII
Means, Standard Deviations, and 95% Confidence Intervals of Pairwise Differences for Place Attachment (PAI) Score for the Recency Factor

Recency group	*N*	*M*	*SD*	Never visited	Visited within last year
Never visited	145	48.63	8.46		
Visited within the last year	44	40.39	9.01	5.29 to 11.21*	
Visited more than one year ago	193	44.73	8.88	2.02 to 5.79*	-7.21 to –1.47*

Note. An asterisk indicates that the 95% confidence interval does not contain zero, and therefore the difference in means is significant at the 0.05 level using the LSD procedure.

Place meaning (YPMS score) versus proximity of residence to Grand Canyon

A one-way analysis of variance was conducted to evaluate the hypothesis that the place meanings Grand Canyon holds for a student would be richer, on average, the closer that student lives to Grand Canyon. The independent variable, the proximity factor, comprised the three levels described above: less than 3 hours driving time, 3—6 hours driving time, and more than 6 hours driving time. The dependent variable was the student's YPMS score. The ANOVA was non-significant, $F(2, 362) = 0.10$, $p = 0.90$. Table VIII shows the means and standard deviations for each level of the factor for the total YPMS score.

Table VIII
Means and Standard Deviations for Place Meaning (YPMS) Score for the Proximity Factor

Proximity Group	*N*	*M*	*SD*
Less than 3 hours	90	58.16	14.72
3—6 hours	216	57.56	16.56
More than 6 hours	59	58.48	13.52

Place meaning (YPMS score) versus total number of visits to Grand Canyon

A one-way ANOVA was conducted to test the hypothesis that the place meanings Grand Canyon holds for a student would be richer, on average, the more times that student visits Grand Canyon in total. The independent variable, the total visit frequency factor, included the three levels explained above: never visited, visited one to three times, and visited more than three times. The dependent variable was the student's YPMS score. The ANOVA was significant, $F(2, 370) = 7.08$, $p = 0.001$. The strength of the relationship between the total number of times visiting Grand Canyon and YPMS score, as assessed by η^2, was medium, with the total visit frequency factor accounting for 37% of the variance of the dependent variable.

To evaluate pairwise differences among the means, follow-up tests were again conducted. Because the variances among the three groups ranged from 13.25 to 13.31 it was assumed that the variances were homogeneous, and post-hoc comparisons were made using the LSD test. There were significant differences in the means between all of the groups (all $p < 0.05$). The group that had never visited Grand Canyon showed a lower YPMS score (i.e., Grand Canyon place meanings were less rich or weaker for this group) in comparison to the group that visited one to three times in total, and in comparison to the group that visited more than three times in total. The group that visited a total of one to three times also showed a lower YPMS score in comparison to the group that visited more than three times in total. The 95% confidence intervals for the pairwise differences, as well as the means and standard deviations for each level of the factor, are shown in Table IX.

Table IX
Means, Standard Deviations, and 95% Confidence Intervals of Pairwise Differences for Place Meaning (YPMS) Score for the Total Visit Frequency Factor

Total visit frequency group	*N*	*M*	*SD*	Zero times	1-3 times
Never visited	141	60.93	16.31		
1—3 times	193	57.02	15.03	0.57 to 7.26*	
More than 3 times	39	50.92	13.25	4.54 to 15.47*	0.79 to 11.40*

Note. An asterisk indicates that the 95% confidence interval does not contain zero, and therefore the difference in means is significant at the 0.05 level using the LSD procedure.

Place meaning (YPMS score) versus number of visits to Grand Canyon within the last year

A one-way ANOVA was conducted to test the hypothesis that the place meanings Grand Canyon holds for a student would be richer, on average, the more times that student visited Grand Canyon within the last year. The independent variable, the one-year visit frequency factor, included the three levels discussed above: never visited in the past year, visited one to three times in the past year, and visited more than three times in the past year. The dependent variable was the student's YPMS score. The ANOVA was significant, $F(2, 365) = 6.02$, $p < 0.01$. The strength of the relationship between the number of times visiting Grand Canyon within the last year and YPMS score, as assessed by η^2, was very small, with the one-year visit frequency factor accounting for only 3.2% of the variance of the dependent variable.

Follow-up tests were again conducted to evaluate pairwise differences among the means. Because the variances among the three groups ranged from 12.07 to 15.76 it was assumed that the variances were homogeneous. Post-hoc comparisons were again made using the LSD test. There were significant differences in the means between never visiting Grand Canyon in the last year and visiting one to three times in the last year. There were no significant differences between never visiting in the last year and visiting more than three times in the last year. Neither were there significant differences between visiting one to three times in the last year and visiting more than three times in the last year. The group that had never visited showed a lower YPMS score in comparison to the group that visited one to three times, and in comparison to the group that visited more than three times. The 95% confidence intervals for the pairwise differences as well as the means and standard deviations for each level of the factor are shown in Table X.

Table X
Means, Standard Deviations, and 95% Confidence Intervals of Pairwise Differences for Place Meaning (YPMS) Score for the One-Year Visit Frequency Factor

One-year visit frequency group	*N*	*M*	*SD*	Zero times	1-3 times
Never visited	325	58.85	15.76		
1-3 times	42	50.95	12.07	2.94 to 12.86*	
More than 3 times	1	81.00		-52.46 to 8.17	-60.67 to 0.58

Note. An asterisk indicates that the 95% confidence interval does not contain zero, and therefore the difference in means is significant at the 0.05 level using the LSD procedure.

Place meaning (YPMS score) versus length of time since last visit to Grand Canyon

A one-way ANOVA was conducted to test the hypothesis that the place meanings Grand Canyon holds for a student would be richer, on average, the more recently that student has visited Grand Canyon. The independent variable, the recency factor, included the three levels discussed above: never visited, visited within the last year, and visited more than one year ago. The dependent variable was the student's YPMS score. The ANOVA was significant, $F(2, 366) = 6.52$, $p < 0.01$. The strength of the relationship between how recently someone has visited Grand Canyon and YPMS score, as assessed by η^2, was very small, with the visit frequency factor accounting for only 3.4% of the variance of the dependent variable.

Follow-up tests were then conducted to evaluate pairwise differences among the means. Because the variances among the three groups ranged from 13.21 to 16.35 it was again assumed that the variances were homogeneous; post-hoc comparisons were made using the LSD test. There were significant differences in the means between all groups of length of time since visiting Grand Canyon ($p < 0.05$). The group that has never visited showed a lower YPMS score in comparison to the group that visited within the last year, and in comparison to the group that visited more than one year ago. The group that visited within the last year showed a higher YPMS score in comparison to the group that had visited more than one year ago. The 95% confidence intervals for the pairwise differences, as well as the means and standard deviations for each level of the factor, are shown in Table XI.

Table XI
Means, Standard Deviations, and 95% Confidence Intervals of Pairwise Differences for Place Meaning (YPMS) Score for the Recency Factor

Recency group	*N*	*M*	*SD*	Never visited	Visited within last year
Never visited	139	61.07	16.35		
Visited within the last year	46	52.07	13.21	3.86 to 14.15*	
Visited more than one year ago	184	57.08	15.13	0.60 to 7.40*	-10.00 to -.0024*

Note. An asterisk indicates that the 95% confidence interval does not contain zero, and therefore the difference in means is significant at the 0.05 level using the LSD procedure.

All results are summarized in Table XII.

Table XII
Summary of Relationships between Proximity and Visitation Factors and Student's Sense of the Study Place (Grand Canyon)

Factor	Does this factor significantly affect	
	Place attachment?	Place meaning?
Proximity of Residence	No	No
Total Number of Visits	Yes (More visits = Stronger attachment)	Yes (More visits = Richer meaning)
Number of Visits in the Last Year	Yes (More frequent visits = Stronger attachment)	Yes (More frequent visits = Richer meaning)
Length of Time Since Last Visit	Yes (More recent visits = Stronger attachment)	Yes (More recent visits = Richer meaning)

Discussion

In this study, proximity and visitation factors that reflect the level of prior experience with Grand Canyon, suggested by previous research to be related to sense of place, were compared to measurements of place attachment and place meaning in order to determine whether these factors have any influence on student's sense of the study place prior to the place-based intervention.

Proximity of Grand Canyon to the places where the geology students live or have lived does not appear to have any influence on their prior senses of the place. This result may simply reflect unfamiliarity with regional geography, as there was no way to confirm the accuracy of student responses to the question of distance from their homes to Grand Canyon. However, it may also confirm the point discussed above, that living close to a place could just as readily provoke indifference ("one doesn't go camping in one's backyard"), boredom, or negativity (Pretty, Chipuer, & Bramston, 2003) as affirmative place attachment.

However, both emotional attachment to and richness of meaning represented by Grand Canyon were positively correlated with the frequency and recency of visits there. This result is concordant with the tourism-related findings discussed above (Williams, Patterson, Roggenbuck, & Watson, 1992), and further confirms that individuals are more likely to make repeat visits to places they value and enjoy; that experiences at the actual Grand Canyon are richer and more meaningful than those imparted remotely by videos, images, or writings; and that the affective and cognitive effects of experiences at Grand Canyon will be strongest in those who have visited it the most recently.

As discussed above, Grand Canyon was selected as the subject of this study because of its recognizability and broad familiarity. It was assumed that these would enhance effect. It is certainly possible that not all of the student respondents who were familiar with Grand Canyon had a positive association with the place. However, the positive correlation between visitation and place attachment suggests that any negative contribution from place aversion was minimal.

It should also be noted that perception of the content encoded in the items of the PAI and especially the YPMS is highly subjective, and the numerical scales of these instruments may be understood somewhat differently by respondents and the researcher (Vázquez, Manassero, & Acevedo, 2006). Future sense-of-place instruments could be made more valid by enabling respondents to express a level of agreement with different statements pertaining to meanings of a place (rather than words or short phrases as in the YPMS), each reviewed and scaled beforehand by a panel of recognized experts on that place (Vázquez, Manassero, & Acevedo, 2006; Semken & Butler Freeman, 2008).

A practical implication of these findings for place-based geoscience teaching, which is consciously intended to leverage and enhance the sense of place (Semken, 2005; Lim & Calabrese Barton, 2006), is that an instructor need not be concerned that the method will be effective only for locally resident students, particularly when the study place or places are widely known and richly imbued with humanistic as well as scientific meaning. Ideally, however, all students should be afforded opportunities to visit and explore these places if it is at all practical. In designing and implementing a place-based geoscience course or curriculum, the instructor should be broadly aware of students' interests, preferences, and prior experiences related to regional travel and outdoor activities.

Acknowledgments

The authors heartily thank the lead instructor, chief teaching assistant, and regular graduate teaching assistants in the spring 2005 physical-geology laboratory course for their assistance and patience. We are also indebted to the 400-plus students in the course for their willingness to participate in this study.

References

Aikenhead, G. S. (1997). Toward a First Nations cross-cultural science and technology curriculum. *Science Education, 81,* 217-238.

Aikenhead, G. (2001). Integrating Western and Aboriginal sciences: Cross-cultural science teaching. *Research in Science Education, 31,* 337-355.

Aikenhead, G. S., Calabrese Barton, A., & Chinn, P. W. U. (2006). Forum: Toward a politics of place-based science education. *Cultural Studies of Science Education, 1,* 403-416.

Aikenhead, G. S., & Jegede, O. J. (1999). Cross-cultural science education: a cognitive explanation of a cultural phenomenon. *Journal of Research in Science Teaching, 36,* 269-287.

Aikenhead, G. S., & Ryan, A. G. (1992). The development of a new instrument: "Views on Science-Technology-Society" (VOSTS). *Science Education, 76,* 477-491.

Athman, J., & Monroe, M. (2004). The effects of environment-based education on students' achievement motivation. *Journal of Interpretation Research, 9,* 9-25.

Ault, C. R. (2008). Achieving *querencia:* Integrating a sense of place with disciplined thinking. *Curriculum Inquiry, 38,* 605-637.

Barab, S. A., & Roth, W-M. (2006). Curriculum-based ecosystems: Supporting knowing from an ecological perspective. *Educational Researcher, 35,* 3-13.

Beus, S. S., & Morales, M. (Eds.). (2003). *Grand Canyon geology (2nd ed.).* New York: Oxford University Press.

Brandenburg, A. M., & Carroll, M. S. (1995). Your place or mine?: the effect of place creation on environmental values and landscape meanings. *Society and Natural Resources, 8,* 381-398.

Bricker, K. S., & Kerstetter, D. L. (2002). An interpretation of special place meanings whitewater recreationists attach to the South Fork of the American River. *Tourism Geographies, 4,* 396-425.

Brown, B. B. (1987). Territoriality. In D. Stokols & I. Altman (Eds.), *Handbook of environmental psychology* (pp. 505-531). New York: John Wiley & Sons.

Butler, R. F., Hall-Wallace, M., & Burgess, T. (2000). A sense of place: At home with local natural history. *Journal of College Science Teaching, 30*, 252-255.

Cajete, G. (2000). *Native science: natural laws of interdependence.* Santa Fé, NM: Clear Light Publishers.

Casey, E. S. (1996). How to get from space to place in a fairly short stretch of time: Phenomenological prolegomena. In S. Feld & K. H. Basso (Eds.), *Senses of place* (pp. 13-52). Santa Fé, NM: School of American Research Press.

Chinn, P. W. U. (2006). Preparing science teachers for culturally diverse students: developing cultural literacy through cultural immersion, cultural translators, and communities of practice. *Cultural Studies of Science Education, 1,* 367-402.

Clark, J. K., & Stein, T. V. (2003). Incorporating the natural landscape within an assessment of community attachment. *Forest Science, 49,* 867-876.

Emekauwa, E. (2004). *The star with my name: The Alaska Rural Systemic Initiative and the impact of place-based education on Native student achievement.* Arlington, VA: The Rural School and Community Trust.

Ernst, J. A., & Monroe, M. (2004). The effects of environment-based education on students' critical thinking skills and disposition toward critical thinking. *Environmental Education Research, 10,* 507-522.

Glasson, G. E., Frykholm, J. A., Mhango, N. A., & Phiri, A. D. (2006). Understanding the Earth systems of Malawi: Ecological sustainability, culture, and place-based education. *Science Education, 90,* 660-680.

Gruenewald, D. A. (2003a). Foundations of place: a multidisciplinary framework for place-conscious education. *American Educational Research Journal, 40,* 619-654.

Gruenewald, D. A. (2003b). The best of both worlds: A critical pedagogy of place. *Educational Researcher, 32,* 3-12.

Gruenewald, D. A., & Smith, G. A. (Eds.). (2008). *Place-based education in the global age: Local diversity.* New York: Lawrence Erlbaum Associates.

Hirst, S. (2006). *I am the Grand Canyon: The story of the Havasupai people.* Grand Canyon, AZ: Grand Canyon Association.

Kaltenborn, B. P. (1998). Effects of sense of place on responses to environmental impacts: a study among residents in Svalbard in the Norwegian high Arctic. *Applied Geography, 18,* 169-189.

Kawagley, A. O., Norris-Tull, D., & Norris-Tull, R. A. (1998). The Indigenous worldview of Yupiaq culture: its scientific nature and relevance to the practice and teaching of science. *Journal of Research in Science Teaching, 35,* 133-144.

Korpela, K. M. (1989). Place-identity as a product of environmental self-regulation. *Journal of Environmental Psychology, 9,* 241-256.

Levine, R., González, R., Cole, S., Fuhrman, M., & Le Floch, K. C. (2007). The geoscience pipeline: a conceptual framework. *Journal of Geoscience Education, 55,* 458-468.

Lieberman, G. A., & Hoody, L. L. (1998). *Closing the achievement gap: Using the environment as an integrating context for learning.* San Diego, CA: State Education and Environment Roundtable.

Lim, M., & Calabrese Barton, A. (2006). Science learning and a sense of place in a urban middle school. *Cultural Studies of Science Education, 1,* 107-142.

Mishler, E. G. (1990). Validation in inquiry-guided research: The role of exemplars in narrative studies. *Harvard Educational Review, 60,* 415-442.

Moore, R. L., & Graefe, A. R. (1994). Attachment to recreation settings: The case of rail-trail users. *Leisure Sciences, 16,* 17-31.

Morehouse, B. J. (1996). *A place called Grand Canyon: Contested geographies.* Tucson, AZ: Universityof Arizona Press.

Powell, J. W. (1987). *The exploration of the Colorado River and its canyons.* New York: Penguin. (Original work published 1895)

Powers, A. L. (2004). An evaluation of four place-based education programs. *Journal of Environmental Education, 35,* 17-32.

Pretty, G. H., Chipuer, H. M., & Bramston, P. (2003). Sense of place amongst adolescents and adults in two rural Australian towns: The discriminating features of place attachment, sense of community, and place dependence in relation to place identity. *Journal of Environmental Psychology, 23,* 273-287.

Proshansky, H. M. (1978). The city and self-identity. *Environment and Behavior, 10,* 147-169.

Proshansky, H. M., Fabian, A. K., & Kaminoff, R. (1983). Place-identity: Physical world socialization of the self. *Journal of Environmental Psychology, 3,* 57-83.

Pyne, S. J. (1998). *How the Canyon became Grand: A short history.* New York: Penguin Books.

Relph, E. (1976). *Place and placelessness.* London: Pion Limited.

Reynolds, S. J., Johnson, J. K., & Stump, E. (2005). *Observing and interpreting geology: a laboratory manual for introductory geology* (pp. 111-120). Tucson, AZ: Terra Chroma.

Riggs, E. M. (2005). Field-based education and indigenous knowledge: Essential components of geoscience education for Native American communities. *Science Education, 89,* 296-313.

Ryden, K. C. (1993). *Mapping the invisible landscape: Folklore, writing, and the sense of place*. Iowa City, IA: University of Iowa Press.

Semken, S. (2005). Sense of place and place-based introductory geoscience teaching for American Indian and Alaska Native undergraduates. *Journal of Geoscience Education,* 53, 149-157.

Semken, S., & Butler Freeman, C. (2008). Sense of place in the practice and assessment of place-based science teaching. *Science Education, 92,* 1042-1057.

Semken, S. C., & Morgan, F. (1997). Navajo pedagogy and Earth systems. *Journal of Geoscience Education, 45,* 109-112.

Shamai, S. (1991). Sense of place: an empirical measurement. *Geoforum, 22,* 347-358.

Smith, G. A. (2002). Place-based education: Learning to be where we are. *Phi Delta Kappan, 8,* 584-594.

Sobel, D. (2004). *Place-based education: connecting classrooms and communities.* Great Barrington, MA: The Orion Society.

Stokols, D., & Shumaker, S. A. (1981). People in places: A transactional view of settings. In J. H. Harvey (Ed.), *Cognition, social behavior, and the environment* (pp. 441-488). Hillsdale, NJ: L. Erlbaum.

Tuan, Y-F. (1977). *Space and place: the perspective of experience*. Minneapolis, MN: University of Minnesota Press.

Vázquez, A., Manassero, M. A., & Acevedo, J. A. (2006). An analysis of complex multiple-choice science-technology-society items: Methodological development and preliminary results. *Science Education, 90,* 681-706.

Williams, D. R., Patterson, M. E., Roggenbuck, J. W., & Watson, A. E. (1992). Beyond the commodity metaphor: Examining emotional and symbolic attachment to place. *Leisure Sciences, 14,* 29-46.

Williams, D. R., & Stewart, S. I. (1998). Sense of place: an elusive concept that is finding a home in ecosystem management. *Journal of Forestry, 96,* 18-23.

Williams, D. R., & Vaske, J. J. (2003). The measurement of place attachment: validity and generalizability of a psychometric approach. *Forest Science, 49,* 830-840.

Wohlgschaft, H. (1994). *Grosse Karl May biographie: Leben und werk.* Paderborn, Germany: Igel Verlag Wissenschaft.

Woodhouse, J. L., & Knapp, C. E. (2000). *Place-based curriculum and instruction: Outdoor and environmental education approaches* (Digest EDO-RC-00-6). Charleston, WV: ERIC Clearinghouse on Rural Education and Small Schools, Appalachia Educational Laboratory. (ERIC Document Reproduction Service No. ED448012)

Young, M. (1999). The social construction of tourist places. *Australian Geographer, 30,* 373-389.

Concept Maps as Tools for Assessing Students' Epistemologies of Science

Emily J. Borda
Western Washington University

Donald J. Burgess
Western Washington University

Charlotte J. Plog
Western Washington University

Natalia C. DeKalb
Western Washington University

Morgan M. Luce
Western Washington University

Abstract

The use of concept maps as instruments for assessing preservice teachers' epistemologies of science (their ideas of the nature of scientific knowledge) was evaluated in this study. Twenty-three preservice elementary teachers' responses to the Views of the Nature Of Science (VNOS) questionnaire were compared to concept maps created in response to the general probe, "What is science?" While VNOS responses allowed a richer analysis of the content and quality of the participants' epistemologies, the concept maps provided information about structural changes of participants' epistemologies as well as how those epistemologies relate to their overall conceptions of science as a field of study. Both instruments also revealed important connections between NOS tenets, which were more numerous on the concept maps but more informative on the VNOS, and between NOS tenets and pedagogical issues. Implications for assessment of students' epistemologies of science in classrooms are discussed.

Correspondence should be addressed to Emily J. Borda, Phone: (360)650-3135; Fax: (360)650-2826; bordae@wwu.edu; 516 High St MS 9150, Bellingham, WA 98225

Introduction

It is widely agreed that scientific literacy involves not just understanding scientific ideas, but understanding the nature of scientific knowledge, or having an informed epistemology of science (AAAS, 1990, 1993; National Research Council, 1996). Thus, a host of education reform efforts include improved instruction assessment aimed at helping students adopt sophisticated ideas about the structure and function of scientific knowledge (Abd-El-Khalick & Lederman, 2000; Lederman, 1992; Meichtry, 1993). Here we describe the application of concept maps, a common assessment tool, toward the assessment of students' epistemologies of science.

Epistemology of Science

The term epistemology is defined differently in different bodies of literature. Epistemology was first conceived as a branch of philosophy concerned with the nature of knowledge and knowing. Psychologists use the term slightly differently, often referring to the term personal epistemology as a *students' beliefs about* the nature of knowledge and knowing. One of the first such lines of research was William Perry's (1970) longitudinal study of college students which resulted in the development of a scheme for characterizing students' epistemologies. According to this scheme, individuals move from ideas about knowledge as certain, unproblematic, and either wrong or right, through a radical relativist phase in which all knowledge is seen as equally valid due to its tentativeness, and finally come to recognize that while knowledge is inherently uncertain, the merits of competing claims can and should be evaluated based on a non-arbitrary set of standards.

Although Perry's research was not discipline specific, it laid a foundation for classifying students' epistemologies of science. Carey, Evans, Honda, Jay, and Unger (1989), for example, used data from interviews with 7th graders to develop a parallel scheme, at the lowest level of which scientific knowledge is viewed as being unproblematic and "read" directly from nature. Students progress toward more sophisticated ideas about theories being mental constructs relying on human interpretation of indirect evidence and finally, at the highest level, recognize that theories must explain all of the available evidence and therefore sometimes must change to accommodate new evidence. Other researchers (Elby, 2001; Hofer & Pintrich, 1997) have used epistemology as an overarching term for a students' beliefs about *learning*. Elby (2001) developed an instrument to measure students' "epistemological beliefs," part of which asks students about the relative merits of conceptual vs. algorithmic learning.

The phrase nature of science (NOS) is closely related to the philosophical treatment of the term epistemology, applied specifically to the realm of science. It is widely used to refer to the nature and function of scientific knowledge (Abd-El-Khalick, Bell, & Lederman, 1998; Lederman, 1992). What constitutes a sophisticated understanding of NOS, or epistemology of science? Scientific developments in the 20th century convinced philosophers to reinterpret strict logical positivism which claims that science can uncover objective "truths" in nature. While most philosophers of science now recognize science as a constructive endeavor in which human interpretation plays a necessary role, many students still have a positivist view of science. Although there is some disagreement among scientists and philosophers of science about what, exactly, constitutes a sophisticated epistemology of science, Smith and colleagues argue that such disagreements are irrelevant to elementary, secondary, and perhaps even tertiary levels of instruction, and that sufficient consensus exists among such scholars to define a relatively robust set of NOS learning goals for schools (M. U. Smith, Lederman, Bell, McComas, & Clough, 1997). By studying consensus views among philosophers of science, Lederman and colleagues identified a set of tenets that, together, constitute a sophisticated epistemology of science. These are: a) evidence forms the basis of scientific theories (empirical NOS); b) there is no single method that automatically generates scientifically valid knowledge (myth of scientific method); c) the practice of interpreting evidence is infused with the backgrounds, expertise and values of the scientist (theory-laden NOS); d) science and society are tightly intertwined and influence each other (social/cultural embededness); e) most scientific practices require the scientist to exercise creativity and imagination (creative and imaginative NOS); f) scientific theories are subject to change based on the accumulation of new evidence and re-interpretations of existing evidence (tentative NOS); g) theories and laws are different kinds of knowledge and one does not become the other (theories vs. laws); and h) the construction of scientific theories requires interpretation of evidence (Lederman, Abd-El-Khalick, Bell, & Schwartz, 2002; Lederman & O'Mally, 1990).

These tenets are consistent with the developmental frameworks described above, wherein a sophisticated epistemology involves recognition of the tentativeness of knowledge due to the roles of interpretation and human interaction in generating that knowledge. In fact, Akerson, Buzzelli, and Donnelly (2007) found that students who had naïve conceptions of NOS relative to Lederman's tenets were generally found to be at the lower levels of Perry's (1970) scheme while those who had more sophisticated conceptions were generally higher on that scheme. Furthermore, the NOS tenets echo the national standards for NOS learning goals set out by the national research council (National Research Council, 1996). The NOS framework will be used to define a sophisticated epistemology in this study, where the word epistemology is used in the philosophical sense to mean the nature of knowledge and knowing, specifically tied to scientific knowledge.

Assessment of Students' Epistemologies

Research has shown that college undergraduates consistently hold, and sometimes leave college with, naïve epistemologies. Perry's findings, for example, suggested that most students do not attain the most sophisticated levels of his epistemological scheme by the time they graduate from college (Perry, 1970). Smith and Wenk (2006) found that the highest epistemological level in Carey's (1989) scheme reached by a group of 35 freshmen was level 2 (3 is the highest), and only one third of the students reached that level. Several studies conducted within Lederman's NOS framework suggest college students and preservice teachers hold inadequate conceptions of the nature of science (Abd-El-Khalick, 2005; Abd-El-Khalick & Akerson, 2004; Lederman, 1992), and that even those who gain sophistication in their NOS conceptions have difficulty retaining and applying them (V. Akerson, Morrison, J. and McDuffie, A., 2006). Further, Abd-El-Khalick (2001) has shown that explicit NOS instruction can result in students adopting "anything goes" epistemologies similar to Perry's naïve-relativistic developmental phase.

A first step in designing instructional strategies geared toward fostering sophisticated epistemologies is to effectively assess students' incoming ideas about science. Here we operate under the framework that a sophisticated epistemology of science involves understanding multiple facets of the nature of science in a coherent, connected way. We therefore use Lederman's NOS framework to define and assess a students' epistemologies because it allows for the measurement of each of these dimensions, rather than placing students on a single spectrum as in Perry's (1970) and Carey's (1989) schemes.

A number of Likert and multiple choice instruments have been developed to assess students' epistemologies throughout the years (e.g. Billeh & Hasan, 1975; Cooley & Klopfer, 1961; Cotham & Smith, 1981). However, these types of instruments assume students interpret the questions and statements in the same way the instructor or researcher using the instrument does, which is not likely to be true. Furthermore, responses to such instruments usually only reveal to what degree a students' views agree with the researchers', rather than giving a holistic view of the students' epistemologies. Recognizing such limitations, Lederman and colleagues developed an open-ended instrument called the Views of the Nature Of Science (VNOS) questionnaire (Lederman et al., 2002). The VNOS questionnaire, when paired with appropriate interview strategies, can give the researcher or instructor more nuanced insights into a student's ideas regarding the eight tenets compared to multiple choice or Likert-type instruments. However, even when combined with interview data, the VNOS requires a great deal of inference on the researcher's part to make sense of the responses.

A limitation of many assessment tools is they do not adequately probe the connections students make between topics. It is well documented that "experts" in a realm chunk ideas together in meaningful ways instead of filing individual pieces of information away to recall piecemeal (e.g. Miller, 1956). We argue that a sophisticated epistemology should therefore include not only nuanced views with respect to individual tenets, but also a recognition of how different aspects of science are related to each other. However, few studies have explored the links students make between different facets of their epistemologies of science. Southerland, Johnston, Sowell, and Settlage (2005) constructed conceptual ecologies representing five graduate students' epistemologies of science based on multiple data sources. They identified common links between facets of the nature of science and suggested a certain hierarchy of ideas, wherein gaining understanding of certain ideas would facilitate development of overall sophisticated epistemologies more than understanding certain other tenets. Schwartz, Lederman, and Crawford (2004) recognized the importance of making connections between NOS tenets, looking for "a demonstrated shift from viewing aspects of NOS as separate components to realizing the interrelationships of the aspects" (p. 625) as evidence for sophisticated epistemologies. These researchers did uncover some connections, illustrating another advantage of such open-ended questionnaires over forced-choice instruments. Nevertheless, we wondered to what extent concept mapping, which explicitly requires students to make connections between different facets of their understanding, could complement the VNOS.

Concept Mapping

Concept mapping has been in existence for more than two decades (J. D. Novak & Gowin, 1984). At their most basic level, concept maps consist of a number of concepts related to a topic, connected to each other via links in a hierarchical or web-based form. In many cases, the links themselves are labeled to describe in words the relationship they represent. Concept maps have been used widely as a tool to assess student understanding (Edmondson, 2000). They are somewhat unique among assessment tools because in addition to assessing the *quality* of the student's understanding (through the number and relationships between concepts and examples), they make the *structure* of that understanding transparent (through the way in which the concept map is organized; Mintzes, Wandersee, & Novak, 2001; Joseph D. Novak, 1984). Importantly, these two facets (quality and structure) seem to be related. In one study, comparison of students' interview responses with the their concept maps revealed that the students who drew more complex maps had a deeper understanding of the content – that is, the structure of a student's understanding seemed to be positively correlated with the sophistication of his or her understanding (Markham & Mintzes, 1994).

Researchers have created several ways to use concept maps to assess student understanding. Novak (1984) describes three important facets of a concept map: 1) propositions, which consist of pairs of concepts connected by linking words, the quantity and validity of which are related to the quality of the individual's understanding of the topic; 2) the levels of hierarchy in a map, which are related to the extent to which the individual subsumes, or groups, more specific knowledge under more general knowledge; and 3) crosslinks, or links between concepts in different branches of the map, which are evidence of knowledge integration, or the extent to which the individual recognizes the connectedness of the ideas within a topic. In a later framework, the number of concepts and examples were added to the facets of a concept map that portray the quality of an individual's understanding, and branching was defined as being reflective of the degree of knowledge differentiation, or the extent to which specific components of concepts are identified (Markham & Mintzes, 1994).

Concept mapping has also been used to reveal the extent of reorganization of a student's knowledge structure. Rummelhart and Norman (1978) have described three types of changes in the way individuals reorganize knowledge: accretion, in which new knowledge is added to existing knowledge structures, tuning, in which the accuracy of the knowledge structure is changed, and reconstruction, in which new knowledge structures replace the old ones. In a similar vein, Carey (1987) called the addition of new knowledge into an existing knowledge structure weak conceptual change and a reconstruction of the knowledge structure strong conceptual change. Jones and Vesilind (1994) used these ideas to assess concept maps drawn by preservice teachers. They defined the set of concepts connected to the central, or level-one concept in a concept map, superordinate concepts. Changes in superordinate concepts indicated the extent to which the teachers reconstructed their conceptual frameworks.

Here we describe a study in which concept mapping was used to assess students' epistemologies of science. The use of concept mapping in this study was similar to that described by Spector, Strong, and La Porta (1998), in which students constructed and modified concept maps about the nature of science throughout the duration of a course. In that study, concept mapping was used primarily as a learning tool. However, to our knowledge the effectiveness of concept mapping as a tool to assess students' epistemologies has not been explicitly investigated. To this end, we undertook a mixed-methods approach in which we compared students' concept maps about the nature of science to their responses on the VNOS questionnaire. The guiding questions for this study were, What are the strengths and weaknesses of concept mapping as a tool for assessing preservice teachers' epistemologies of science? and, What new capabilities can concept mapping offer for assessing preservice teachers' epistemologies of science?

Method

In this study, preservice elementary teachers in one section of a science methods course responded to the VNOS questionnaire and engaged in a concept mapping activity both before and after instruction in the nature of science. The responses to both instruments were compared to gain more understanding of how concept mapping functions as a tool to assess students' epistemologies of science.

Context of study

The second author was the instructor for the science methods course in which this study was conducted. Developing sophisticated epistemologies of science is a major goal of this course. Inquiry based activities (Lederman & Abd-El-Khalick, 1998) and readings (AAAS, 1990, 1993) were used to explicitly teach NOS ideas. All twenty-three study participants were enrolled in the same section of the methods course and were pursuing a K-8 certificate.

Data collection

All data collection and analysis was carried out by the first, third, fourth and fifth authors. All participants responded to selected items from the VNOS-C questionnaire (Lederman et al., 2002) before and after instruction in the nature of science. Eight students who held a broad range of views were then chosen to participate in semi-structured interviews to ensure the validity of researchers' interpretations participants' written responses. All interviews were audio recorded and transcribed for analysis.

Near the beginning of the course, participants were instructed in concept mapping and generated practice concept maps in groups of 3-4. The participants then brainstormed words or phrases related to the following questions: What is science? What is the scientific world view? What is scientific inquiry? Who does science and how do they do it? The groups generated concept maps based on the brainstormed ideas. After instruction in the nature of science, groups were given back their initial posters and instructed to make changes by adding, moving or removing concepts and links. After each concept mapping task, the groups sketched their maps and turned them in to the researchers for analysis.

Data analysis

Data from the VNOS questionnaire and concept mapping activity were compared in three phases (Table 1). In phase I we used each instrument to normatively assess the quality of each participant's epistemology. We then explored the participants' epistemologies descriptively in phase II. Finally, in phase III explored to what extent each instrument could give information about the structure, or connectedness, of participants' epistemologies.

Table 1
Description of data analysis activities for each instrument in each phase of data analysis.

	Purpose	VNOS	Concept maps
Phase I	Normative (quality of students' epistemologies)	Naïve/informed coding of each NOS tenet	Scoring based on the number of concepts, valid relationships and examples
Phase II	Descriptive (nature of students' epistemologies)	Generation and coding of emergent themes	Generation and coding of emergent themes
Phase III	Structural (structure of students' epistemologies)	Identification of links between NOS tenets;	a) Identification of links between NOS tenets; b) Scoring based on the number of branches, levels of hierarchy and crosslinks; c) Identification of changes in superordinate concepts

Phase I. In the phase I, responses to the VNOS and related interview questions were blinded and coded independently by two authors after at least 90% interrater reliability was reached in training sessions. Each of the participants was coded naïve or informed with respect to the first six of the eight NOS tenets described in the introduction. Interview responses were used to establish validity of the coding procedure. The scoring rubric and interview procedure were adapted from Lederman et al. (2002) and Bell et al. (2005). The codes were then used to generate an overall epistemology profile of "naïve," "emergent" or "informed." If a participant expressed informed understandings of 5-6 of the six tenets, he or she was given an overall code of "informed." For 3-4 informed codes, the participant was given an "emergent" overall code, and "naïve" for 0-2 (Bell et al., 2005). Finally, the VNOS responses of one student who exhibited a large overall shift (naïve to informed) were examined for ideas that seemed key to her epistemological change. The quality of students' epistemologies was judged on their concept maps by the number of *a)* concepts, *b)* examples and *c)* valid relationships (only connections with valid linking terms were counted).

Phase II. Both VNOS responses and concept maps were thoroughly searched for emergent themes in this phase (Strauss & Corbin, 1998). Because the questions asked in the concept mapping exercises (What is science? What is the scientific world view? What is scientific inquiry? Who does science and how do they do it?) were most similar to the first item on the VNOS questionnaire ("What, in your view, is science? What makes science . . . different from other disciplines of inquiry. . .?" (Lederman et al., 2002)), only emergent themes from the first question were used for comparison with the concept maps. Responses from each instrument were first coded individually then searched for themes generated from coding the other instrument, in order to allow a comparison.

Phase III. The final analysis phase consisted of searching the VNOS responses and concept maps for connections between two or more different tenets. The emergent themes from phase two analysis were placed into families based on the six of Lederman's NOS tenets assessed in this study. Using ATLAS.ti, we searched VNOS responses for co-occurrences of two or more tenets in the same statement. Search results were read individually to ensure the links were accurate. We then searched for linkages between the same six NOS tenets on the concept maps. Linker words used to connect two or more concepts related to these tenets served as evidence for how the participants made sense of the relationships between them. Additional measures of structure were taken on the concept maps only. First, the frequency of the structural features of each concept map – branching, hierarchies and crosslinks – were tallied to give a structure subtotal. Second, changes in the second highest level (superordinate) concepts were tracked in order to illuminate the level of conceptual change represented.

Comparing the VNOS responses and concept maps. Because the VNOS questionnaires were completed individually and concept maps were completed in groups, we could only compare whole-class results on the two instruments (ie. we did not compare student 1's VNOS responses with her group's concept map because the latter would represent more views than student 1's). In phase I, we compared the naïve/informed code frequencies from the VNOS to the concept, relationship and example scores on the concept maps, which we combined into a concept subtotal. This comparison allowed us to evaluate how well each instrument assessed the students' understandings of the nature of science. We could not quantitatively compare the code frequencies to the concept subtotals because they are different statistics. However, we are able to discuss what each statistic revealed about students' NOS understandings, as well as the strengths and limitations of each measure. In phase II, we compared the emergent codes created for the responses to the first question of the VNOS and propositions (concepts + linker words which form statements) on the concept maps. The frequency of each code on each group of instruments was tallied and compared. Finally, in phase III, we compared the number and types of links between NOS tenets on each instrument. The structure subtotals and superordinate concepts were features unique to the concept maps. Though we didn't have a means for comparing these features to VNOS responses we felt it important to evaluate their contribution to the overall assessment of students' epistemologies and thus discuss them as stand-alone features.

Results

The findings with respect to each of the three phases of data analysis are described below. In quotations from VNOS responses, participants are identified individually by a number. In quotations from the concept maps, the seven concept mapping groups are assigned letters A-G.

Quality of participants' epistemologies

VNOS. The pre- and post-instruction VNOS codes are shown in Table 2. Modest gains were observed for all but one tenet. None of these gains were found to be statistically significant through a chi-square test in which the distribution of students holding naïve and informed views with respect to each tenet was compared before and after instruction. The overall codes, shown in Table 3, reveal 6 of the 23 participants adopted more informed views as a result of instruction. In contrast, one student moved from an informed to emergent profile after instruction.

Table 2
Pre- and post-instruction NOS beliefs of participants by tenet, as assessed through analysis of responses to the VNOS questionnaire.

	Informed		Naive	
NOS Tenet	Pre	Post	Pre	Post
Creativity & Imagination	9 (39%)	10 (43%)	14 (61%)	13 (57%)
Empirical NOS	4 (17%)	6 (26%)	19 (78%)	17 (74%)
Myth of Scientific method	5 (22%)	4 (17%)	17 (74%)	18 (78%)
Social/cultural embededness	12 (52%)	14 (61%)	10 (43%)	8 (35%)
Tentative NOS	19 (83%)	21 (91%)	4 (17%)	2 (9%)
Theory-laden NOS	11 (48%)	16 (70%)	11 (48%)	7 (30%)

Note. The percentage of naïve and informed responses on each questionnaire do not always total 100% because in some cases we did not find evidence to justify the assignment of either code.

Table 3
Overall NOS profiles before and after instruction, as assessed by the number of tenets coded informed on the VNOS questionnaire.

Overall Profile	Pre	Post	Δ from N	Δ from E	Δ from I
Informed	2 (9%)	5 (22%)	2 (9%)	2 (9%)	N/A
Emergent	9 (39%)	10 (43%)	2 (9%)	N/A	1 (4%)
Naive	12 (52%)	8 (35%)	N/A	0 (0%)	0 (0%)

To illustrate the subtleties that can be revealed by the VNOS and to illustrate some of the coding employed, one participant's epistemology is here described in more depth. This participant (17), given the alias Phoebe, is one of the two who exhibited an overall naïve to informed shift. Phoebe's VNOS responses signified major shifts in her thinking about three tenets: the myth of scientific method, the theory-laden nature of science, and the social-cultural embededness of science. Therefore, the discussion is focused around these three issues.

Phoebe's initial VNOS responses revealed an accurate conception of an experiment as, "a means of showing a cause and effect relationship. Experiments rely on a controlled environment where the researchers manipulate the variables to discover the effect on the outcome" (response to item 2). However, when asked whether experiments are necessary for the development of scientific knowledge (item 3), Phoebe initially said, "Yes, because an experiment is the only way of knowing for sure if there is a cause and effect relationship between two variables." After instruction, Phoebe appeared to broaden her conceptions about scientific investigations, stating,

> I no longer think the development of scientific knowledge always requires an experiment. An experiment is a great and necessary tool for discovering cause and effect relationships. However, observation is another tool that is used to develop scientific knowledge. For example, scientists have discovered much about the universe through observation and rational thought.

Phoebe was initially given a naïve code in the scientific method category because she did not recognize the utility of observational studies. Her later response, however, was given an informed code in the same category. Phoebe's shift in thinking seemed to hinge first on a correct conception of what an experiment is, and secondly on her ability to recognize that there are instances in which experiments are impossible or inappropriate.

Changes in Phoebe's perception of the theory-laden nature of science (that scientists bring their own experiences and values to bear in collecting and interpreting data) seemed tightly related to changes in her thinking about the social and cultural embededness of science. When asked how scientists using the same evidence can come to different conclusions about what caused the extinction of the dinosaurs (item 5), Phoebe responded, "Some scientists might come to one conclusion using the evidence and some another." Her lack of inclusion of a mechanism for how this might happen resulted in a naïve code for the theory-laden category. After instruction, Phoebe revised her answer to this question, stating, "Scientists come from different parts of the world and their culture could also play a role in how they analyze evidence." Her inclusion of the idea that scientists' cultures could play a role in analyzing evidence indicates an understanding of the interpretive role of the scientist, and that this comes in part from the different backgrounds, experiences, and values of the scientists involved in a study. This response provided a basis for informed codes for both the theory-laden category and the socio-cultural category. Furthermore, when asked whether science is universal or infused with certain cultural values (item 6) at the end of the class, Phoebe responded,

> I still believe science is intended to be universal, but cultural values do play a role. For example, during the Middle Ages scientists believed that men carried all the genetic material for their children, this reflected the culture at the time.

This response indicates some understanding of the way a scientific theory may influence cultural norms. Phoebe therefore recognized the interplay between science and a society's culture both in the context of a scientist conducting a study and in the larger sense of a theory influencing a culture. This set of responses formed the basis for an informed code in the socio-cultural category whereas at the beginning of the course Phoebe had a more simplistic view:

> I believe science is universal. Science relies on observation and reproducible data. If scientists in China perform an experiment that shows a new drug is effective in fighting cancer, scientists in other countries must be able to perform the same experiment and get the same result.

Although her statement is not incorrect, it reveals a lack of understanding of the more nuanced interplays between science and the society in which it is embedded.

This narrative illustrates the type of information that can be obtained using the VNOS questionnaire. We were able to ascertain some of the major facets of Phoebe's epistemology that changed as a result of instruction. These included increased understandings of the importance of observational studies in science, the role of interpretation in science, and the interplay between scientific theories and cultural norms.

Concept maps. Sample concept maps are shown in Figures 1 and 2. The attribution of the number of concepts and/or valid relationships in a concept map to the extent of the mapper's understanding of the topic is well established (Liu, 2004; Markham & Mintzes, 1994; Joseph D. Novak, 1984). We therefore used scores generated by counting the number of concepts, valid relationships and examples to obtain information about the extent of participants' understanding of the nature of science (Table 4). We observed a statistically significant gain (using a Wilcoxon t-test for nonparametric data) in the number of valid relationships. We also observed a non-statistically significant gain in the number of concepts represented. The number of examples, however, decreased. The concept subtotal (concept, relationship and examples scores totaled) was higher at the end of instruction than the beginning for each concept map, and the increase in the mean was statistically significant.

Figure 1. Pre-instruction concept map created by group E.

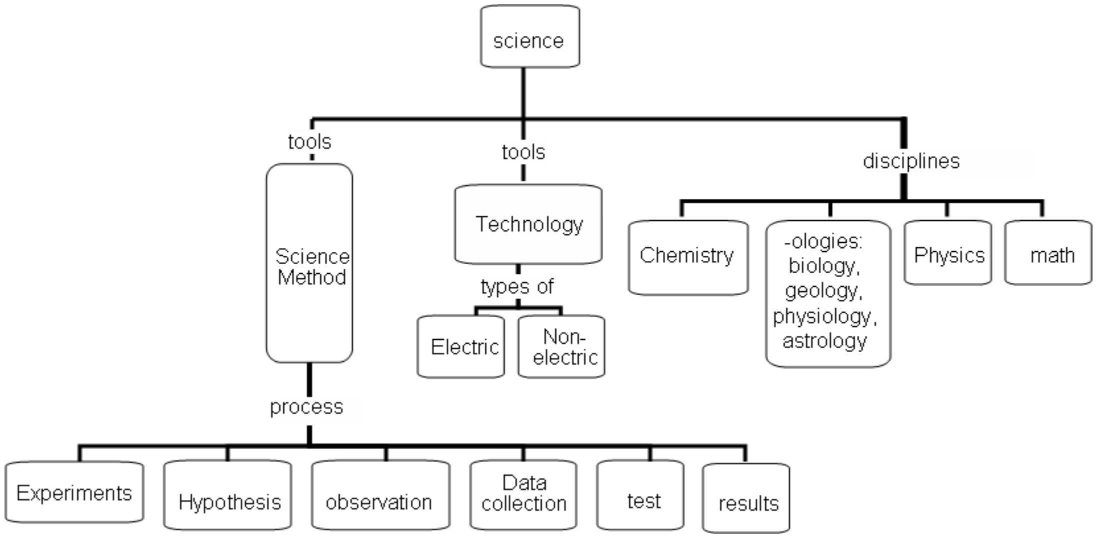


Figure 2. Post-instruction concept map created by group E.

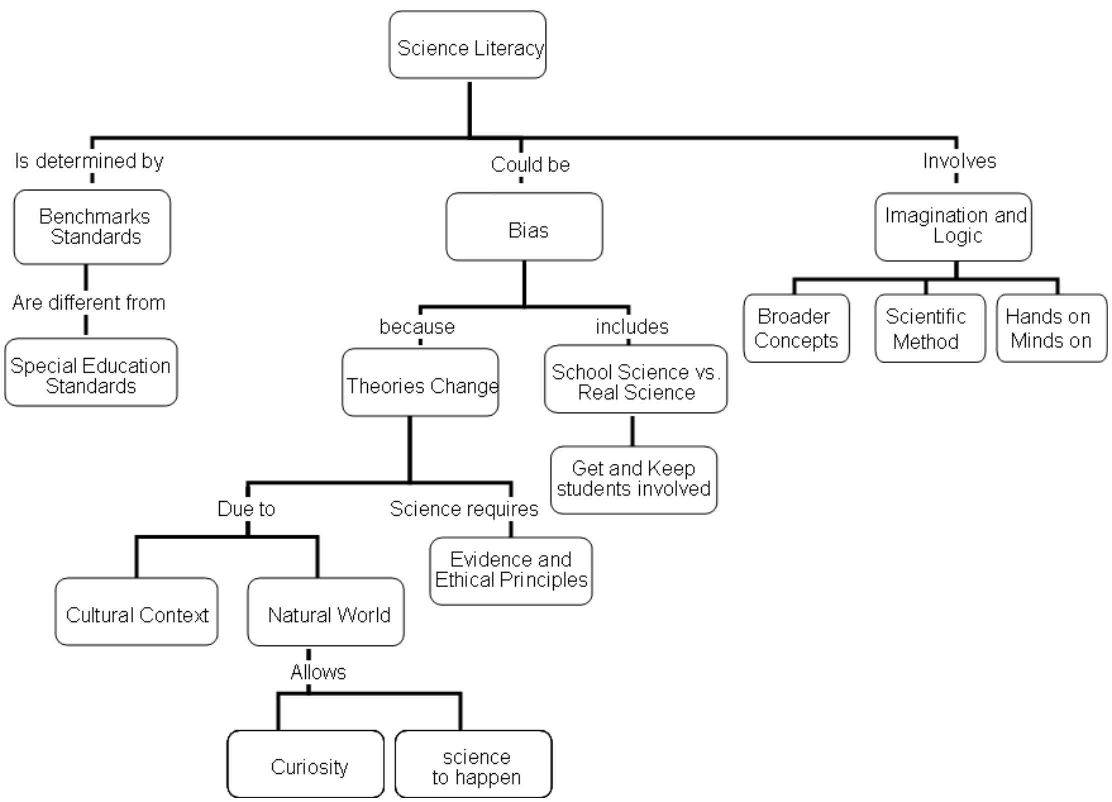


Table 4
Average concept map scores.

	Pre-instruction		Post-instruction	
	Mean score	Std. Dev.	Mean score	Std. Dev.
Concepts	16.9	8.3	27	7.8
Relationships	6.4	4.9	19.6*	8.9
Examples	5.4	4.7	1.9	4.9
Concept Subtotal[a]	**36**	**10.7**	**48.0***	14.9
Branching[b]	13.9	5.7	17.5	5.5
Hierarchies[c]	16.4	5.6	24.3	7.3
Crosslinks[d]	5.7	11.3	15.7	22.3
Structure Subtotal[a]	**36**	**13**	**58.2**	23.1

[a]Subtotals were calculated as means of individual concept map subtotals, not by adding the mean scores for each category. [b]One point was awarded to the first branch, and three points were awarded for every subsequent branch. [c]Five points were awarded for each level of hierarchy. [d]Ten points were awarded for each crosslink (Markham & Mintzes, 1994). $*p < 0.05$

Description of participants' epistemologies

Many themes emerged from the qualitative analysis of both VNOS responses and concept maps. We sorted these themes into five categories: a) the goals of science; b) elements of scientific inquiry; c) elements of scientific knowledge; d) other elements unique to science; and e) elements of science teaching and learning. Frequencies of themes related to each category on the questionnaires and maps are shown in Table 5.

Table 5
Responses to VNOS item 1 and concept map entries referencing one or more themes related to five emergent categories identified in the two instruments.

Category	VNOS item 1		Concept Maps	
	Pre	Post	Pre	Post
Goals of science	15 (65%)	5 (22%)	4 (57%)	2 (29%)
Elements of scientific inquiry	14 (61%)	18 (78%)	6 (86%)	7 (100%)
Elements of scientific knowledge	14 (61%)	7 (30%)	6 (86%)	7 (100%)
Other elements unique to science	0 (0%)	1 (4%)	7 (100%)	2 (29%)
Elements of science teaching	0 (0%)	3 (13%)	0 (0%)	7 (100%)

The goals of science category includes statements or concepts related to the purpose(s) of science. Many participants saw the goal(s) of science as studying the world and/or asking questions. Responses such as, "I believe science is understanding the world and what makes up the world" (20**,** pre-VNOS) were common on the first VNOS item. Themes related to scientific processes were grouped into the elements of scientific inquiry category. In this category, participants commonly referenced the roles of creativity and/or imagination, curiosity, experiments, hands-on activities, inquiry, scientific method, objectivity, observation and reasoning in generating scientific knowledge.

The category elements of scientific knowledge includes themes related to ways in which scientific knowledge is different from other types of knowledge. The most common themes in this category referenced the importance of evidence, facts, hypotheses and theories. Many pre-instruction concept maps and responses to the first VNOS item included references to evidence: "Science, in general, takes empirical evidence to support theories" (1, pre-VNOS). Several post-instruction concept maps and VNOS responses included tentativeness as a defining element of scientific knowledge: "science is different from other inquiry based disciplines because it is ever-changing and uses imagination, logic and evidence to come up with theories" (9, post-VNOS).

On many of the concept maps participants indicated miscellaneous facets of science that, to them, distinguished it from other disciplines. These themes were grouped into the category other elements unique to science. The most frequent of these were entries on the concept maps related to various science disciplines (biology, geology, etc.) and technology. Some concept maps also included references to science materials, such as test tubes or microscopes. Only one response to VNOS item one fell into the other elements category: "Science also goes hand-in-hand with math and technology" (14, post-VNOS).

The final category included themes related to science teaching and learning. Although most responses in this category were found in the concept maps, there were a few responses to VNOS item one that incorporated such ideas, for example, "Science is something that has students and those studying it investigate things through inquiry based curriculums (*sic*) that inquire rather than answer linear questions" (16, post-VNOS).

Structure of participants' epistemologies

In the previous section we describe the evidence related to the quality and descriptions of participants' epistemologies as assessed through the VNOS questionnaire and the concept maps. Here, we describe evidence related to the structure of the participants' epistemologies.

Concept maps. The number of branches, hierarchies and crosslinks found in each concept map are reflective of the degree of knowledge differentiation, subsumption and integration, respectively (Markham & Mintzes, 1994). Mean scores in each category increased from pre- to post-instruction (Table 4), as did, consequently, the structure subtotal. Although none of these gains were found to be statistically significant through a Wilcoxon t-test, we observed a statistically significant ($p < 0.05$) increase in the overall mean total score from 64.7 to 106.0.

We also examined the second-level, or superordinate concepts for evidence of reconstruction of students' epistemologies (Jones & Vesilind, 1994). In all cases most of the superordinate concepts on the post-instruction concept maps were different than those on the pre-instruction maps (Table 6). The most commonly added superordinate concepts were related to teaching and learning, consistent with the descriptive data above.

Table 6
Changes in superordinate concepts on concept maps.

Group	# Superordinate concepts		Superordinate concepts lost		Superordinate concepts gained	
	Pre	Post	Number	Percent	Number	Percent
A	2	4	1	50	3	75
B	3	3	2	67	2	67
C	9	8	9	100	8	100
D	2	4	1	50	3	75
E	6	6	5	83	5	83
F	4	6	3	75	5	83
G	4	1	4	100	1	100

Links between NOS tenets. We searched both the concept maps and responses to all of the questions on the VNOS questionnaire for links between the NOS tenets assessed in this study. Only 6 links were found in the VNOS responses, all but one in post-instruction responses. Links involving some mention of evidence, which we grouped under the empirical NOS tenet, were most common in responses to the VNOS questionnaire. An example such a link is: "Inferences must be made from the evidence, and at times the evidence may be subjective" (19, post-VNOS item 6), in which the empirical NOS is linked to the theory-laden NOS.

Although links between NOS tenets on concept maps were more numerous (3 pre, 17 post) than those found in the VNOS responses, they were not as descriptive because arguments on concept maps consist only of concepts and linker terms. Group D, for example, linked the tentative and empirical NOS tenets in this sequence of concepts and linkers (linker words underlined): "Changing of theories <u>through</u> inquiry <u>demands</u> evidence" (post-concept map).

Links between NOS tenets and ideas related to teaching science. The introduction of concepts related to teaching and learning on all of the post-instruction concept maps provided the opportunity to analyze how the participants' epistemologies were connected to their notions of teaching and learning science. We therefore searched each post-instruction concept map for connections to concepts related to the tenets assessed with the VNOS questionnaire. Three (43%) of the post-instruction concept maps included such connections. Two of these maps included the idea of bias, which was interpreted to be related to the theory-laden nature of science, whether or not the use of this term represented an informed view of this tenet. Group A included the concept "teaching own biases," on their post-instruction maps. Group C connected the idea of tentativeness to inquiry-based education on their post-instruction map: "Views on nature of science <u>includes</u> education <u>with</u> inquiry [which] <u>leads to</u> changing theories." Other links were made between science education, evidence and creativity.

We compared participants' ideas about teaching and learning from the concept maps to those from their responses to the VNOS questionnaire. Only three (13%) participants connected ideas related to teaching and learning science to NOS tenets in their responses to the first VNOS item. One participant referenced the empirical nature of science as an important element in a student's experience: "Students should be allowed to experiment with many things throughout their 'scientific' career, and be allowed to explore evidence on their own rather than always being told by the teacher what is right and wrong" (15, post-VNOS item 2). The idea of bias, interpreted as a facet of the theory-laden nature of science, was cited with reference to teaching by one participant: "Educationally I have found that your opinions of science will be imposed upon the students that you teach. The biases and opinions of every concept will be instilled in the students, including their scientific curiosity" (18, post-instruction VNOS item 5). Finally, one participant expressed the importance of creativity in teaching science: "It is important for teachers to foster imagination and creativity in their students. Students should think of science as fun and exciting, not boring and dogmatic" (17, post-VNOS item 7).

Discussion

The aim of this study was to compare the types of information about students' epistemologies that can be gained using concept maps versus the open-ended VNOS questionnaire, and to assess the strengths and weaknesses of each instrument. To this end, we compared epistemological views of 23 preservice elementary teachers on concept maps to their responses to the VNOS questionnaire before and after instruction in the nature of science. We analyzed the data from both instruments according to three basic considerations: a) quality, b) description and c) structure of participants' epistemologies.

Quality of Participants' Epistemologies

We observed modest gains in the number of informed codes on five of the six NOS tenets assessed on the VNOS. Although none were statistically significant, the concurrent gains in five of the tenets provides initial evidence that participants' epistemologies increased in sophistication. Lack of large gains is consistent with other studies that suggest it is difficult to change students' epistemologies in one course (e.g. (V. Akerson, Morrison, J. and McDuffie, A., 2006)), especially when NOS is not the only focus of a course, as was the case in this study.

Our case-study of Phoebe illustrates another level of normative analysis that is possible with VNOS responses. The set of questions on the VNOS were designed to give a more holistic profile of an individual's epistemology than can be gained through forced-choice instruments (Lederman et al., 2002). We were able to see, for example, that Phoebe's changing ideas about the nature of science hinged upon knowledge about experiments and correlational studies, as well as the ways in which culture and science influence each other, both on a personal level (the individual scientist) and a societal level (theories influencing culture).

The concept maps gave us valuable information about the quality of the participants' epistemologies, but the nature of this information was different from the VNOS responses. We observed a statistically significant gain in the mean concept subtotal score from the concept maps, the main component of which was the number of valid relationships between concepts. Linking words enable students to form propositions or arguments from their chains of concepts (Joseph D. Novak, 1984). Because the number of valid links increased more than the number of concepts or examples (examples, in fact, decreased), we can infer that the largest change in participants' epistemologies was not in their ability to assimilate ideas about NOS piecemeal, but to form new propositions using ideas that were already part of their epistemologies.

We are cautious in comparing the quantitative results from the VNOS and concept maps because they were arrived at differently. While several different concepts and arguments may have contributed to a single VNOS code, each concept, relationship and example was counted individually on the concept maps. However, the purpose of this study was not to establish the validity of either instrument (prior studies support each instrument's validity – see, for example, Lederman et al. (2002) and Markham & Mintzes (1994)), but rather to explore the different types of information each can provide in assessing students' epistemologies of science. Both instruments can be coded in such a way as to provide a quantitative measure level of sophistication of a student's epistemology (number of informed codes on the VNOS vs. concept subtotal on the concept maps). Our data initially suggest these measures are consistent with each other, in that both revealed gains, though we cannot say with confidence that the VNOS gains reflected increased sophistication of participants' epistemologies, as they were not statistically significant. Further studies would have to be done with larger sample sizes to rigorously establish the correlation of the two instruments' quantitative measures. As illustrated by the analysis of Pheobe's responses, the VNOS may also be used to identify key ideas around which students' epistemologies seem to hinge, resulting in a more nuanced picture of the changes that take place when a student goes from a naïve to an informed epistemology. The concept maps, on the other hand, were more useful in identifying to what extent students assimilated new propositions about the nature of science (illustrated by the concepts and examples included on the maps) vs. reworking existing propositions (illustrated by the links).

Descriptions of participants' epistemologies

The VNOS normative codes and concept map scores were augmented with emergent themes analysis of the concept map concepts and VNOS item one responses. Of the five categories into which the emergent themes were grouped, three were related to participants' epistemological commitments: the goals of science, the nature of scientific inquiry and elements of scientific knowledge. In these first three categories there was some degree of consistency between VNOS responses and concepts on the concept maps, in that themes initially identified from one instrument were almost always observed in the other.

The two other categories into which we grouped the themes, other elements unique to science and issues of teaching and learning, were represented differently on the two instruments. Most of the pre-instruction concept maps consisted of entire branches made up of concepts we grouped into other aspects unique to science, such as lists of scientific disciplines and instrumentation. However, neither the pre- nor the post-instruction VNOS responses revealed attempts at defining science as a collection of disciplines or specialized instruments. This distinction may stem from a difference in the level of scaffolding in each task. While the concept mapping task simply required students to map their idea of science, VNOS item one more specifically probes students to think about the difference between science and other disciplines. Perhaps because of the differences in the specificity of the prompts, the concept maps, as they were used in this study, seemed to reveal the relative prominence of a students' epistemology of science in their overall conception of science as a field of study. The VNOS responses, on the other hand, seemed to give more detailed information about the epistemologies themselves.

We attributed differences in representation of themes related to the teaching and learning of science on the two instruments to differences in scaffolding as well. The instruction that occurred between implementation of both instruments was directed towards increasing pre-service teachers' knowledge of science teaching methods in addition to helping them adopt more sophisticated epistemologies. Therefore, it is reasonable to expect students' ideas about teaching methods to be incorporated into their responses to both instruments. While all of the post-instruction concept maps included themes related to teaching and learning science, only 13% of the responses to VNOS item one included such themes. As above, this evidence suggests students felt less constrained by the concept mapping activity and included ideas they may not have seen as relevant to VNOS item one.

In summary, participants used similar ideas on both instruments when describing their epistemologies. The major difference was the degree to which participants included ideas that were not necessarily part of their epistemologies. Because of the less specific nature of the concept mapping task when compared with the more probing VNOS items, we were unable to get as detailed a picture of the participants' epistemologies from their concept maps as we were with their responses to the VNOS questionnaire. However, we were able to get a broader sense for how they think about science – whether it be as a way of knowing, as a collection of disciplines, or as a subject to be learned. This is significant because whether or not a student has a sophisticated epistemology of science may not matter if the student primarily thinks about science as a collection of sub-disciplines instead of as a way of knowing or generating knowledge. Furthermore, the concept maps can lend some insight into how seemingly non-epistemological ideas such as scientific disciplines or teaching and learning practices influence or are influenced by participants' epistemologies of science. Thus, using concept maps we can learn more about how students' epistemologies influence their outlook on science as a discipline and how science should be taught and learned.

Structure of participants' epistemologies

Because of the proposed importance of students' cognitive structures to their ability to understand and apply knowledge (Bransford, Brown, & Cocking, 2000; Markham & Mintzes, 1994), we took several measures of cognitive structure on both the concept maps and the VNOS responses. The degree of branching, number of hierarchies and number of crosslinks factored into the structure subtotal on the concept maps. This subtotal increased for six of the seven post-instruction maps when compared to the pre-instruction maps. The concurrent increases in the structure and concept subtotals support the theory that as the sophistication of a student's understanding increases, so too does the complexity of the structure of his or her understanding (Markham & Mintzes, 1994). The largest gain from the structure sub-categories was in the number of hierarchies used. Since the number of hierarchies represent the extent of knowledge subsumption (Markham & Mintzes, 1994), we can infer that the largest structural changes in participants' epistemologies involved more clustering of specific ideas under overarching ideas, a structuring skill that seems to be positively related to an individual's degree of expertise with a topic (Miller, 1956). Furthermore, we found that 50 – 100% of the superordinate concepts on the post-instruction concept maps were different from those on the pre-instruction maps. Therefore, we conclude that the higher degree of structure on the post-instruction concept maps is indicative of a reorganization of the participants' epistemologies, rather than added levels of organization to an existing core structure (Jones & Vesilind, 1994).

Because none of the structure gains were statistically significant, we cannot dismiss the possibility that these gains were due to chance. The small sample size for the concept maps made it difficult to establish statistical significance, and further studies with individual concept maps are recommended. However, we are able to get a picture for the type of information the concept maps can reveal about students' epistemologies and how that information can be used to supplement VNOS analysis. Concept maps are unique among assessment tools in enabling an instructor or researcher to analyze the degree and type of a students' cognitive reorganization, information which we expect to be helpful in addressing questions related to the process by which students' epistemologies of science change.

In order to gain more descriptive information about some of the structural features of participants' epistemologies, we investigated two different types of connections on participants' concept maps and in their VNOS responses. First, we searched both instruments for connections between the six NOS tenets assessed in this study. Second, we searched both instruments for connections between epistemological ideas and issues related to the teaching and learning of science. Both types of links were found on both instruments to increase after instruction, suggesting students were more able to see interdependencies among the NOS tenets and to integrate these tenets into their conceptions of teaching science as a result of instruction. The increase in NOS-teaching links may have been due to increasingly sophisticated epistemologies, knowledge of teaching practices, and/or the links between them as a result of instruction. We cannot, from our evidence, make a causal claim about this increase. We found that both types of connections were more common on the concept maps than they were in responses to the VNOS questionnaire. This did not surprise us, since concept maps are designed to show links between concepts. However, because participants were limited to linker words in expressing the logic behind their linkages in the concept maps, these were often less informative than connections found in the VNOS responses. Thus, while the concept maps in this case gave us a snapshot of the types of connections students made, the VNOS responses, though fewer, helped us to better understand the nature of some of these connections.

Limitations of this study and suggestions for future research

One limitation of this study is that we did not interview the groups about their concept maps. We recognize some of the concepts and relationships expressed on the concept maps would have been better understood by the researchers if students were asked to explain their concepts and links in an open-ended way. However, we also recognize limitations posed by time and class size make it difficult for some teachers to interview students about concept maps they create. Thus, our findings can shed light on the strengths and weaknesses of concept maps as an assessment tool for epistemologies of science in an authentic classroom context.

Another limitation is that while students answered the VNOS items individually, they participated in the concept mapping activity in groups. Because we cannot be sure to what extent each individual's views were represented on the concept maps, we were unable to do pairwise comparisons. We would therefore suggest that a future study include individual concept mapping exercises, so that such comparisons can be done.

A final limitation is that any time two or more instruments are used on the same sample, there is a chance that one of the instruments could be acting as a treatment in addition to its role as an assessment tool. Indeed, research has suggested concept mapping can be used as a learning tool as well as an assessment instrument (Joseph D. Novak, 1998; Prezler, 2004). Such a treatment effect could be responsible for some of the similarities between responses to the two instruments that were cited above, especially considering that they were administered within a week of each other both before and after instruction. However, we did uncover some important differences between the two instruments, such as the prevalence of concepts related to science disciplines on the concept maps compared with their absence on the VNOS, so we believe this treatment effect was minimal. The extent to which concept maps can help students develop their epistemologies of science is nonetheless a rich topic for further exploration.

Conclusion

Because of the small sample size used in this study we are careful not to generalize our findings to a larger population of preservice teachers. Instead, we wish to use these findings to suggest potential roles for concept mapping in assessing students' epistemologies of science. One important finding from this study is that the concept maps, while giving us less specific information about facets of students' epistemologies of science (their ideas about the nature of scientific knowledge) compared to the VNOS, gave us information about how those epistemologies were situated within their overall ideas about science. One potential hybrid use of the two instruments would therefore be to use concept maps to probe students' initial conceptions of science without leading them too much into a certain framework of thinking about science. The VNOS can then be used as a follow-up tool, either as a written questionnaire, an interview, or both, to provide more information about the quality of students' epistemologies.

We also found that because of the less structured nature of the concept mapping probe, students more often included their ideas about teaching and learning science on their concept maps then they did in their VNOS responses. Concept maps can therefore be used in methods courses to probe how preservice teachers' epistemologies of science relate to their ideas about teaching science. Such an approach would be especially powerful if the preservice teachers were interviewed about the nature of the links between epistemology concepts and teaching concepts.

Finally, we found that the concept maps provided useful information about the structure of participants' epistemologies that the VNOS questionnaire did not. To what extent the participants' epistemological ideas were integrated, differentiated and/or subsumed under other concepts can be much more easily accessed through concept maps than through written responses to a questionnaire. Furthermore, the superordinate concepts can be used to assess the extent of a student's cognitive reorganization, which in this case is related to reorganization of his or her epistemology. Although students' responses on the VNOS didn't provide such detailed structural information, they did provide a richer, more detailed look at their epistemologies. Thus, concept maps, as used in this study, gave detailed information about the types of structural changes in a student's epistemology while the VNOS responses gave us more insight into the nature of those changes. Both instruments revealed important strengths and weaknesses and, when combined, can provide a powerful assessment of students' epistemologies.

References

American Association for the Advancement of Science (AAAS). (1990). *Science for all Americans*. New York: Oxford University Press.

American Association for the Advancement of Science (AAAS). (1993). *Benchmarks for Science Literacy*. New York: Oxford University Press.

Abd-El-Khalick, F. (2001). *Embedding nature of science instruction in preservice elementary science courses: Abandoning scientism but. . .* Journal of Science Teacher Education, 12(3), 215-233.

Abd-El-Khalick, F. (2005). *Developing deeper understandings of nature of science: The impact of a philosophy of science course on preservice science teachers' views and instructional planning*. International Journal of Science Education, 21(1), 15-42.

Abd-El-Khalick, F., & Akerson, V. L. (2004). *Learning about nature of science as conceptual change: Factors that mediate the development of preservice elementary teachers' views of nature of science*. Science Education, 88(5), 785-810.

Abd-El-Khalick, F., Bell, R. L., & Lederman, N. G. (1998). *The nature of science and instructional practice: Making the unnatural natural*. Science Education, 37(6), 563-581.

Abd-El-Khalick, F., & Lederman, N. G. (2000). *Improving science teachers' conceptions of nature of science: A critical review of the literature*. International Journal of Science Education, 22(7), 665-701.

Akerson, V., Morrison, J. and McDuffie, A. (2006). *One Course is Not Enough: Preservice Elementary Teachers' Retention of Improved Views of nature of Science*. Journal of Research in Science Teaching, 43(2), 194-213.

Akerson, V. L., Buzzelli, C. A., & Donnelly, L. A. (2007). *Early childhood teachers' views of nature of science: The influence of intellectual levels, cultural values, and explicit reflective teaching*. Journal of Research in Science Teaching, 45(6), 748-770.

Billeh, V. Y., & Hasan, O. E. (1975). *Factors influencing teachers' gain in understanding the nature of science*. Journal of Research in Science Teaching, 12, 209-219.

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

Carey, S. (1987). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

Carey, S., Evans, R., Honda, M., Jay, E., & Unger, C. (1989). *'An experiment is when you try it and see if it works': A study of grade 7 students' understanding of the construction of scientific knowledge*. International Journal of Science Education, 11, 514-529.

Cooley, W., & Klopfer, L. (1961). *Test on understanding science (Form W)*. Princeton, NJ: Educational Testing Service.

Cotham, J., & Smith, E. (1981). *Development and validation of the conceptions of scientific theories test*. Journal of Research in Science Teaching, 18, 387-396.

Edmondson. (2000). *Assessing science understanding through concept maps*. In Assessing Science Understanding: Academic press.

Elby, A. (2001). *Helping physics students learn how to learn*. Physics Education Research, 69(7), S54-S64.

Hofer, B. K., & Pintrich, P. R. (1997). *The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning*. Review of Educational Research, 67(1), 88-140.

Jones, M. G., & Vesilind, E. (1994, April 4-8). *Changes in the structure of pedagogical knowledge of middle school preservice teachers*. Paper presented at the International Meeting of the American Educational Research Association (AERA), New Orleans.

Lederman, N. G. (1992). *Students' and teachers' conceptions of the nature of science: A review of the research*. Journal of Research in Science Teaching, 29, 331-359.

Lederman, N. G., & Abd-El-Khalick, F. (1998). *The nature of science in science education: Rationales and strategies*. The Netherlands: Kluwer Academic Publishers.

Lederman, N. G., Abd-El-Khalick, F., Bell, R. L., & Schwartz, R. S. (2002). *Views of Nature of Science Questionnaire: Toward valid and meaningful assessment of learners' conceptions of nature of science*. Journal of Research in Science Teaching, 39, 497-521.

Lederman, N. G., & O'Mally, M. (1990). *Students' perceptions of tentativeness in science: Development, use, and sources of change*. Science Education, 74, 225-239.

Liu, X. (2004). *Using concept mapping for assessing and promoting relational conceptual change in science*. Science Education, 88(3), 373-396.

Markham, K., & Mintzes, J. J. (1994). *The concept map as a research and evaluation tool: Further evidence of validity*. Journal of Research in Science Teaching, 31(1), 91-101.

Meichtry, Y. J. (1993). *The impact of science curricula on student views about the nature of science*. Journal of Research in Science Teaching, 30(5), 429-443.

Miller, G. A. (1956). *The magical number seven, plus or minus two. Some limits on our capacity to process information*. Psychological Review, 63, 81-87.

Mintzes, J. J., Wandersee, J. H., & Novak, J. D. (2001). *Assessing understanding in biology*. Journal of Biological Education, 35(3), 118-124.

National Research Council (1996). *National Science Education Standards*. Washington DC: National Academy Press.

Novak, J. D. (1984). *Application of advances in learning theory and philosophy of science to the improvement of chemistry teaching*. Journal of Chemical Education, 61(7), 607-612.

Novak, J. D. (1998). *Learning, Creating, and Using Knowledge: Concept Maps as Facilitative Tools in Schools and Corporations*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. New York: Cambridge University Press.

Perry, W. (1970). *Forms of intellectual and ethical development in the college years: A scheme*. New York: Holt, Rinehart & Winston.

Prezler, R. (2004). *Cooperative conept mapping*. Journal of College Science Teaching, 33(6), 30-35.

Rummelhart, D., & Norman, D. (1978). *Accretion, tuning and restructuring: Three modes of learning*. In Cotton, J. & Klateky, R. (Eds.), Semantic factors in congnition (pp. 37-53). Hillsdale, NJ: Erlbaum.

Schwartz, R. S., Lederman, N. G., & Crawford, B. A. (2004). *Developing views of nature of science in an authentic context: An explicit approach to bridging the gap between nature of science and scientific inquiry*. Science Teacher Education, 88(4), 610-645.

Smith, C. L., & Wenk, L. (2006). *Relations among three aspects of first-year college students' epistemologies of science*. Journal of Research in Science Teaching, 43(8), 747-785.

Smith, M. U., Lederman, N. G., Bell, R. L., McComas, W. F., & Clough, M. P. (1997). *How great is the disagreement about the nature of science? A response to Alters*. Journal of Research in Science Teaching, 34, 1101-1104.

Southerland, S. A., Johnston, A., Sowell, S., & Settlage, J. (2005). *Reinvoking conceptual ecologies: Inservice teachers' conceptual change in nature of science*. Paper presented at the International meeting of the National Association for Research in Science Teaching (NARST), Dallas, TX.

Spector, B., Strong, P., & La Porta, T. (1998). *Teaching the nature of science as an element of science, technology and society*. In McComas, W. F. (Ed.), The Nature of Science in Science Education: Rationales and Strategies (pp. 267-276). Dordrecht: Kluwer Academic Publishers.

Strauss, A., & Corbin, J. (1998). *Basics of Qualitative Research* (2 ed.). Thousand Oaks, CA: SAGE Publications, Inc.

A Comparison of Student Learning in STS vs Those in Directed Inquiry Classes

Robert E. Yager
University of Iowa

Aeran Choi
Kent State University

Stuart O. Yager
Bethel College

Hakan Akcay
Marmara Universitesi

Abstract

Fifteen experienced grade 5-10 teachers each taught two sections of students – one with an STS approach and one following closely the curriculum with a "directed inquiry" approach. Data were collected from five teaching and assessment domains from the two classes. These include: science concepts, science process skills; creativity, attitudes, applications of concept and processes in new contexts. There was no difference found in assessing in the concept domain. In all the other four domains student outcomes were significantly higher for students in the STS sections.

Correspondence should be addressed to Dr. Robert E. Yager, Science Education Center, Room 769 VAN, The University of Iowa, Iowa City, Iowa 52242, 319-335-1189, Robert-yager@uiowa.edu

Introduction

There has been little actual reform in American science education for the past several decades. Educational policies and new science courses and programs have recommended significant changes, but actual classroom practices have not changed. It is apparent that the practice and theory of reform do not coincide (Bybee, 1991; Abell & Lederman, 2007; Weiss, 1993, 1994, Tillotson, 2005). When teacher beliefs are incompatible with the philosophy of science education reform, a gap develops between the intended and the implemented principles of reform (Levitt 2002). Certainly, the teaching advocated by the National Science Education Standards (NRC, 1996) and that advocated in Science for All Americans (AAAS, 1990) have both identified the need for different forms of teaching. Both indicate that the focus on teaching and learning of science must go far beyond the simple transmittal of scientific facts, concepts, and process skills directly to students. But, the interpretations of the needed changes often result in continuation of the status quo in actual classrooms.

Interestingly, inquiry (a new reform focus proposed in the late 50s) is now accepted as a desired focus for teaching as well as a form of content in the NSES. A follow-up publication from NRC entitled "Inquiry and the NSES" elaborates more focus on inquiry and indicates that there are four levels of student inquiry and five distinct features (NRC, 2000, p. 29). These same levels vary from very "open" inquiry to very "directed" inquiry. Too many are quick to adapt "directed" inquiry with little more than superficial changes with no firsthand experiences with students doing their own inquiry! Instruction described in most state standards and most textbooks remain directive. To STS enthusiasts, inquiry is uniquely student-centered and centered on problem situations identified by students. (Some science educators maintain that open-inquiry – student inquiry – is not really possible (Abell & Lederman, 2007). This study provides data for comparing student outcomes in highly student-centered classrooms (i.e., often open inquiry) that characterizes the teaching central to Science/Technology/Society (STS) (NSTA, 2008-09, p. 242) reforms with student outcomes in classrooms that are taught by the same teacher in a much more directed inquiry fashion.

Much previous research has reported solely on teacher perceptions of their own implementation of recommended teaching practices associated with specific reforms (Bybee & Bonnstetter, 1985; Mitchener &Anderson, 1989; Rubba, 1989; Tillotson, 1996, 2005; Luft, 2001; Luft, Roehrig, Patterson, 2003). The majority of teachers in these studies displayed positive perceptions of their teaching and supported the idea of the use of real world contexts for their teaching. Few studies have focused exclusively on teacher use of the teaching strategies that characterize the nine "more emphasis" conditions recommended in the National Science Education Standards (NSES) – (NRC, 1996, p. 52) as specific ways teaching should change. These nine changes indicate that teachers should: 1) understand and respond to individual student's interests, strengths, experiences, and needs; 2) select and adapt curriculum; 3) focus on student understanding and use of scientific knowledge, ideas, and inquiry processes; 4) guide students in active and extended scientific inquiry; 5) provide opportunities for scientific discussion and debate among students; 6) continuously assess student understanding; 7) share responsibility for learning with students; 8) support a classroom community where cooperation, shared responsibility, and mutual respect occurs; 9) work with other teachers to enhance the entire science program (NRC, 1996).

Science-Technology-Society efforts were underway in the U.S. by 1980 and superseded the National Science Teachers Association (1996) and Project 2061 (AAAS, 1990) with its inclusion as a form of science for Project Synthesis funded by NSF in 1978 (Harms & Yager, 1981). After many extensive efforts to implement STS programs, NSTA appointed a Task Force to define STS. This work resulted in a Position Paper unanimously approved by the NSTA Board of Directors in 1990 after four years of debate. STS was defined in the official NSTA position as "the teaching and learning of science <u>and technology</u> in the context of human experiences. Eleven features of STS were identified in the Position Statement to describe needed change in teaching. These essential features characterizing STS include: 1) student identification of problems with local interest and impact; 2) the use of local resources (human and material) to locate information with can be used in problem resolution; 3) the active involvement of students in seeking information that can be applied to solve real-life problems; 4) the extension of learning beyond the class period, the classroom, the school; 5) a focus on the impact of science and technology on individual students; 6) a view that science content is more than concepts which exist for students to master on tests; 7) an emphasis upon process skills which students can use in their own problem resolution; 8) an emphasis upon career awareness – especially careers related to science and technology; 9) opportunities for students to experience citizenship roles as they attempt to resolve societal issues they have identified; 10) identification of ways that science and technology are likely to impact the future; 11) some autonomy in the learning process as individual issues are identified and used as the basis for science study (NSTA, 2008-09, p. 242). This study is an examination of learning outcomes for students enrolled in STS classrooms versus those following a curriculum closely and using mostly "directed" inquiry.

But there is little evidence that the teaching approaches urged by NSES or those characterizing STS are being employed in schools generally (Mitman, Mergendoller, Marchman & Parker, 1987; Rubba, 1989; Weiss, 1993, 1994). The level of success with STS depends on several factors, such as prior experiences of teachers with STS, the level of "inquiry" they are willing and able to try, their attitudes, the extent of cooperation and communication with their colleagues, and the level to which their instruction focuses on student constructions of concepts (Wilsman, 1991; Williams, 1994). Massenzio (2001) has reported that most teachers are not implementing STS approaches because they are very familiar and comfortable with traditional approaches to science teaching. Most teachers still retain the beliefs that learning occurs as a result of direct teaching. This means that most teachers and state mandated reforms are merely a matter of transmitting what they know and what textbooks and other materials include. Mitchener and Anderson (1989) identified five factors that influence teacher perceptions that keep them from implementing an STS approach. These include: concerns over the dilution of science content, discomfort with cooperative learning, difficulty assessing student work, frustrations regarding varying student ability levels, traditional conceptions of the role of the teacher, and unwillingness to deal with issues not part of their own science preparation. Unfortunately, there are few teachers who have learned science with an STS approach as part of their own preparation. One exception can be found in Iowa where there have been continuous efforts and financial support for moves to STS teaching since 1983 and the Project Synthesis conception of needed changes and the research supporting them.

Features of Iowa Chautauqua and the Scope, Sequence, and Coordination Projects as Sponsored by NSTA

The Iowa Chautauqua Program (like the later SS&C effort involving all science teachers in the 20 Iowa middle school districts) has emphasized constructivist teaching practices with an STS philosophy of learning in classrooms. The SS&C project sought to energize <u>all</u> science teachers in the twenty participating districts in the effort. It was one of six state efforts that comprised the NSTA Scope, Sequence, and Coordination Project. The STS efforts associated with SS&C were funded in Iowa with three major grants totaling over four million dollars over a seven year period (1990-1997) not including local funds from industry and other local support. One major feature of Iowa SS&C was the fact that it followed six years of NSF funding ($2.5 million) as the Iowa Chautauqua

Project which was first funded in 1983 and involved interested teachers from across the state. Both projects utilized successful teachers as instructional "partners" who were actively involved with reforms in Iowa where the model was developed and used. Basically these teacher partners were identified because of their understanding of science and reform teaching pedagogies which characterized the "Desired State" of Harms' Project Synthesis (1977). The first report of this major research effort was first published as a part of the NSTA "What Research Says" series (Harms & Yager, 1981).

The Iowa Chautauqua Model provided the framework for working with Iowa teachers and their implementation of SS&C during the 1990-97 funding interim and continued for four more years with State funding. The Iowa effort provided the framework for the Staff Development efforts for both the funded Chautauqua series and for the Iowa SS&C project. Both were ultimately validated as exemplary by the U.S. Department of Education's Program Effectiveness Panel, precursor for funding by the National Diffusion Network. The major facets of the professional development projects included a leadership conference, three-week summer workshop, 5 to 10 day trial use of the materials and approaches planned during the summer, a 3-day fall short course, interim interactions with other teachers in the study, and a Spring 3-day short course to report on all the efforts for the entire academic year. Some of the goals and products typically collected and evaluated at each stage are also indicated in Figure I. All of the efforts were offered with six teaching and assessment domains central to STS teaching and learning. The STS efforts were found to be superior to more traditional textbook-dominated courses in all domains – sometimes, however, no significant advantages were found for the STS in the Concept Domain (Yager & Tamir, 1992; Yager & Weld, 1999).

Previous studies of the Iowa SS&C Project pertaining to student achievement show that all aspects of the learning of students in STS/Constructivist classrooms increase, especially related to process skills, thinking and designing skills, achievement test scores, ability to apply concepts and skills to new situations, creativity skills, and the development of more positive attitudes concerning science, science study, and science careers (Yager & Weld, 1999; Yager & Tamir, 1992; Harms &Yager, 1981; Yager, 1982; Yager, 1993; Yager & Yager, 2007; Varrella, 1997; Enger, 1997; Yutakom, 1997). This means that typical content (organized around major science concepts) is but one form of content and often the only one aspect used traditionally to assess learning. A broader view of content was developed in Iowa where teachers helped urge such language and focus for the National Science Education Standards (NSES) which were conceived and funded in 1992; they were published in final form in 1996 (NRC, 1996).

Leadership Conference

A Two Week Long Conference Designed To

1. Prepare staff teams for conducting a workshop series which enrolled up to 30 new teachers.
    a) One lead teacher per ten new teachers
    b) Scientists from a variety of disciplines
    c) Scientists from industry
    d) School Administrators
    e) Science Supervisors/Coordinators
2. Organization and scheduling for each workshop
3. Publicity and reporting
4. Assessment strategies
    a) Five domains for assessing students for teaching effectiveness
    b) Use of past reports and sample instruments and techniques
    c) Action Research (Every teacher as researcher)
    d) New research plans for the successful teachers that were instructional partners

Three Week Summer Workshop

Learning Experiences

1. Includes special activities and field experiences that relate specific content within the disciplines of biology, chemistry, earth science, and physics.
2. Makes connections between science, technology, society within the context of real world issues and in terms of meeting the four goals elaborated in the NSES, p. 13..
3. Issues such as air quality, water quality, land use/management are used as the contexts for concept and process skills development.
4. Focuses on problems/issues in the school and local communities.
5. Enrollees develop materials for use in peer teaching as well as specific plans for teaching a 5-10 day mini-module prior to the fall short course.
6. Decisions regarding specific evidence needed to assure that each goal was achieved.

Academic Year Workshop Series

Fall Short Course → (3 days) | Interim Projects → | Spring Short Course (3 days)

Awareness Workshop

<u>20 hr Instructional Block</u>
(Thursday pm. Friday, & Saturday)

Activities Include:

1. Review problems with traditional views of science and science teaching
2. Outline essence of new instructional strategies
3. Define techniques for developing new modules and assessing their effectiveness
4. Select a tentative module topic
5. Practice with specific assessment tools in each Domain.
6. Use Lesson Study designs
7. Analyze one videotape of one class prepared for use in the Short course to be Shared with total group

Three Month Interim Projects

<u>Developing More Modules</u>

Activities Include:

1. Developing instructional plans for minimum of twenty days
2. Administer pre-tests in six domains
3. Teach one complete module (3-4 weeks)
4. Collect posttest information
5. Communicate with regional staff, Partner Teachers, and central Chautauqua staff
6. Complete and analyze one class videotape with colleagues from given sites
7. Decide on other modules to be tried

Final Workshop

<u>20 hr Instructional Block</u>
(Thursday pm. Friday, & Saturday)

Activities Include:

1. Report on new instructional experiences
2. Report on all assessment efforts
3. Interact with new information concerning the new teaching strategies elaborated in the NSES, p. 52
4. Show and discuss one videotape of teaching in one class
5. Analyze changes from summer, fall, and spring
6. Plan for involvement in professional meetings
7. Plan for next-step initiatives (including complete reorganizing of existing courses and helping with new workshop series)

Figure I. The IOWA CHAUTAQUA MODEL: A Professional Development Model Approved by the National Diffusion Network

How SS&C Relates to the Visions Included in the National Science Education Standards

The NSES include eight facets of content including: 1) unifying concepts and processes; 2) science as inquiry; 3) physical science; 4) life science; 5) earth /space science; 6) science and technology; 7) science for meeting personal and societal challenges; and 8) history and nature of science. Unfortunately, many state standards, and most textbooks, focus only on the concept facets (i.e., physical, life, earth/space) and on inquiry (too often at the "teacher directed level") while the other four facets are not commonly pursued by teachers or even acknowledged in some state standards. The Iowa reform efforts have focused on all eight facets of content in addition to concern for attitude and creativity, which have been called the enabling domains.

Assessment efforts in Iowa have utilized the design proposed by Wiggins and McTighe (1998) called "Understanding by Design" which calls for teachers and staff development leaders to develop protocols for assessing evidence that the specific goals advanced have been met before considering any curriculum materials. The stated goals for K-12 science included in the NSES include only four for determining student learning, namely students should: 1) experience the richness and excitement of knowing about and understanding the natural world; 2) use appropriate scientific processes and principles in making personal decisions; 3) engage intelligently in public discourse and debate about matters of scientific and technological concern; and 4) increase their economic productivity through the use of the knowledge, understanding, and skills of the scientifically literate person in their careers. (NRC, 1996, p. 13)

These goals are very similar to the Project Synthesis goals used in the 1980 research project mentioned earlier. Again, there were but four goals. These Synthesis four are: 1) science as preparation for further study; 2) science for dealing with personal problems; 3) science for resolving current societal issues; and 4) science for preparing for science careers (NSTA, Vol. 3 of What Research Says to the Science Teacher). The Synthesis research team reported that the only goal that was identified to plan the curriculum and used to justify science programs was the first one, i.e., preparing students for further study of science across grade levels and for college. It is extremely interesting to note that this goal was omitted completely from the NSES, especially since it was the only goal acclaimed by 95% of the K-12 science teachers. In its place was the new first goal for science in the NSES which indicated that all students must Do Science. Some argue that this first goal is the over-arching one which makes inquiry basic – and a form of content as well as a teaching approach. This is certainly central to STS efforts and to the Iowa SS&C efforts over the seven years it was funded and continues today with Title IIa funding.

With the faculty and focus of the Iowa Staff development efforts during the funding, 1983-1997, it is interesting to see what occurred in the classrooms of the most successful teachers of Iowa SS&C who served as important partners in the continuing professional development efforts. Further, it is of interest to determine what they do in their classrooms that can affect learning in multiple domains.

Changes in Constructivist Teaching Practices

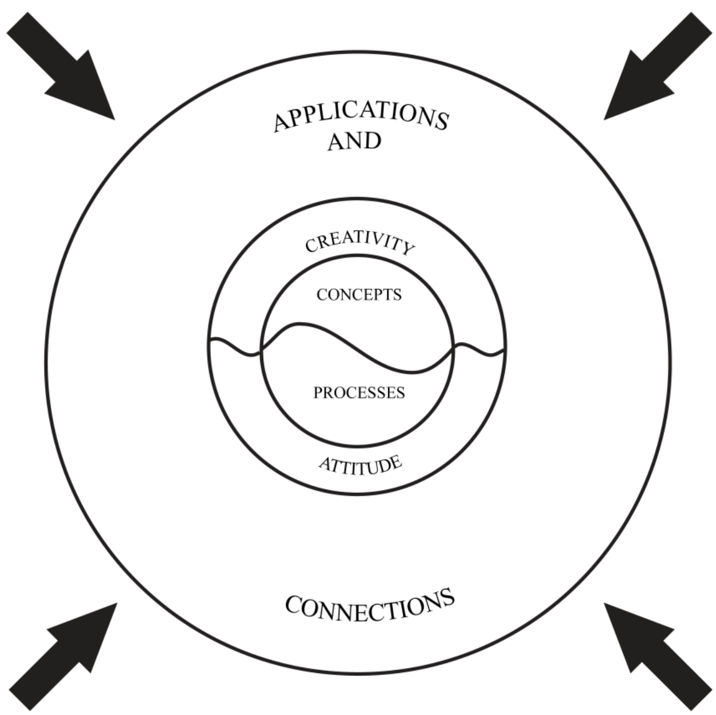
The study of teacher perceptions, teaching practices, and the relationship between them is considered important for understanding new problem teaching situations and for encouraging even more successes with the current reform efforts (Anderson & Mitchener, 1994). The thinking by teachers about their own teaching as well as their implementation of innovative teaching methods provide a basis for understanding the process for accomplishing even more changes. Clark and Peterson (1986) have asserted that teacher behavior is influenced substantially by the thought processes of teachers. Studies of teaching practices are important for enhancing student understanding and improving teaching (Good & Brophy, 2000; Rosenshine, 1971). More importantly, such studies support the fact that effective teaching practices can be developed when teachers are provided with appropriate experiences as a part of continuing staff development projects (Yager & Penick, 1990; Yeany & Padilla, 1986; Kimble, 1999). Teaching practices exhibited by successful teachers serve as guides and inspiration for other teachers to emulate.

Yager and McCormack (1989) have identified six domains for use in science teaching and assessment of teaching success that correspond to the changes advocated in the NSES. These were developed further and used with both Iowa Chautauqua and Iowa SS&C. They were used by the teacher leaders as well as new teachers enrolled in subsequent years. Obviously the teacher leaders provided great role models and the most significant results with new students over subsequent years. These domains include: 1) Concept Domain: mastering basic content constructs; 2) Science Process Domain: learning and using the skills scientists use in "sciencing"; 3) Creativity Domain: improving in quantity and quality of questions, possible explanations, and predicted consequences; 4) Attitude Domain: developing more positive feelings concerning the usefulness of science, science study, science teachers, and science careers; 5) Application and Connection Domain: using concepts and processes in new situations; 6) Worldview Domain: focusing on the whole science enterprise for learning with respect to philosophy, history, and sociology.

Figure 2 illustrates the relationship of the domains identified for assessing successes with the STS approach. The typical domains are concept and process mastery – often the primary foci for typical instruction -- almost always giving more attention to the mastery of given concepts. STS demands attention to the two "enabling" domains that surround the "bulls eye" of the diagram. Biologists like to think of creativity and attitude as symbolizing the cell membrane -- controlling what gets in and out of the world of professional scientists, estimated to be the 0.004 percent of the population of the world who are practicing research scientists. The large Applications and Connections domain is where most people live and work – where the concepts and processes can affect their living. This domain is even larger than indicated in Figure II. Few who even start in college science courses actually operate in the central region (i.e., new concepts and process skills). The sixth domain was not a focus for this study (the Worldview Domain) – where less focus was given to it. Further, the attention to it varied and many instruments that were used were inappropriate for the grade level span. Interest and focus among the instruments and the data collection in the Worldview domain varied widely – both in terms of quantity and quality.

It is also important to emphasize that STS focuses on technology (the design world) as well as on pure science. It connects the world of science to the whole of society. STS efforts also illustrate the importance of society and its affect on people, including scientists. Science and even more so – technology -- focus on human problems and their possible resolution when STS approaches are used. In spite of the call for unification of science and technology, relatively few mergers have occurred in the U.S. nor in Iowa in spite of the major STS focus.

Specific research questions for the study include how students taught by the same teacher in an STS section compare to students in a Non-STS section related to: 1) learning of basic science concepts? 2) learning of science process skills? 3) learning of creativity skills? 4) student attitudes concerning science? 5) ability of students to apply concepts and process skills in new situations? In all fifteen schools each teacher taught one section with an STS approach and one which did not utilize any of the defining characteristics of STS. Hence the results reported in the tables indicate the differences found in STS and Non-STS sections. All 30 classes (15 STS and 15 Non-STS) were taught involving teachers who had been active partners in the Iowa Chautauqua and Iowa SS&C projects – both with major NSF funding. The study involved a total of 310 students in the 15 STS classes and 302 in the Non-STS sections.



Arrows indicate views of science by philosophers, historians, and sociologists

Worldview

Figure II. Domains for Teaching and Assessing Science Learning

Procedures

This is a report of the use of STS teaching approaches and its effects on student learning in five of the six teaching and assessment domains indicated. Fifteen teachers were identified as the most effective partners involved with STS reforms over the thirty year interim. These teachers served as instructional partners for at least ten years. Most were selected from a follow-up study conducted by Yutakom (1997) —all were recommended by the program staff, school administrators, enrolled teachers, and students who provided feedback that illustrated the advantages of STS teaching. It was part of the information used to gain approval from the Program Effectiveness Panel (PEP, 1992)) and ultimately to gain recognition and funding as part of the National Diffusion Network.

Each of the 15 teachers agreed to select one class session where STS would be implemented fully and one class where traditional procedures would be utilized but recognizing the importance of inquiry. The science topics for both sections were the same from the various Standards and science curricula developed and used in the fifteen schools and involved teachers across grades 5 through 10. The differences were the degrees of understanding of the basic science and technology content. In the STS sections students were asked to identify problem questions that framed their study; in the traditional section teachers merely outlined the content that would frame the studies without input from students. Data were collected over two 9-week grading periods which occurred as a part of a whole semester in grades 5-9 middle schools where the selected fifteen teachers taught across Iowa. Data consisted of weekly quizzes as well as unit and semester exams for noting differences in the concept domain.

Other instruments consisted of a process skills instrument used with SS&C and published in "Assessing Student Understanding in Science" (Enger & Yager, 2001). It was given as a pre-test prior to beginning the semester long study and again at the end of the semester.

The attitude instrument was taken from the affective battery characterizing the 1978 administration of National Assessment of Educational Progress (NAEP, 1978) which was the first year that such items were included other than those assessing Concept Mastery. Thirty items were selected where students were asked to indicate their feelings using a scale of: strongly agree, agree, not sure, disagree, strongly disagree. This scale was used as a pre-test before instruction and again at the end of the semester long course.

The creativity scale was used and outlined in the Iowa Assessment Handbook (Yager, Kellerman, & Blunck, 1992). It consists of describing a discrepant event and then recording student questions, suggestions for possible answers, evidence for validity, and predictions of possible consequences, and finally the uniqueness of each on a ten point scale. Each was evaluated in terms of quantity and later for degree of uniqueness. A five point scale was used and rated as follows: response is "irrelevant", i.e., not related to the question (0 score), "pertinent" question related but not creative (2 points), and "unique" difficult to see the connection and not frequently cited by others (5 points). The entire exercise was undertaken as a pre-test, repeated at the end of each unit over the semester, and again as the post-test for the study. Two research assistants (some included as co-authors of this report) reviewed the information – mainly from sample videotapes of at least three class periods at the end of each nine week assessment.

The instruments for assessing student growth in each of the domains are all illustrative of the samples from Enger & Yager (2009). No specific measures were used in the Worldview Domain for all teachers (hence no data are reported or used by all teachers in all thirty classrooms involved with this study). There were consistent directions for using each instrument. Information about validity and reliability issues are reported in Enger & Yager (2009). These domains were defined as a goal of science education in the Iowa Chautauqua Program. The reliability coefficients for assessment items in each of the domains are obtained by using the test-retest method with students in classes taught by all lead teachers for a given year. Specific information reported for this study did not involve the teachers and students in the sample. In other Action Research efforts instrument reliability and/or validity were studied. The reliability regarding the domains ranged from 0.76-0.96 (test-retest two weeks later).

The specific instruments for the five assessment domains are summarized as follows: 1) Concept – The pre-test for the content for the instructional units for the semester and as a final each 9 weeks and an end-of-semester grading). These examples were different for each teacher and grade level. 2) Process – The process test included in Enger and Yager used as a pre-instructional test and a final semester measure. 3) Creativity – Pre and post scores on a discrepant event where students were asked to provide questions, possible-explanations, evidence for the validity of an expert, and an indication of uniqueness for each. Points were given for number and relative quality of each of the four areas. 4) Attitude – Thirty items from the 1978 NAEP were selected and used as a pre-instruction and semester end instrument. 5) Applications – The teachers were assisted in providing application items for each major concept from each unit – these were completed prior to and following each instructional unit and as a final semester survey. Teachers were also active in urging use of the skills and concepts taught in new situations as a feature of STS teaching and an activity to encourage use of the information in ways students could evaluate their own work and that of others in the class. These procedures were great in illustrating assessment as a basic ingredient of science itself – not something that only teachers do to grade student performances. The application domain consisted of items concerning each of the concepts encountered where students were asked to apply them in new situations. Some of these were used to indicate the degree of creativity as well. The "Assessing Student Understanding" monograph (Enger & Yager 2001) details how teachers were asked to provide application items to measure a sampling of content in each of the instructional units and to get experience with use of a multiple choice format. Teachers provided the researchers with examples of their concept items and matched application items for each instructional unit included for each nine-week grading period (often three units involved with each grading period). Teachers in STS sections frequently asked students to suggest their own applications in the classroom, the school, and the wider community.

Results

Tables I through V include the pre and post scores for students in both the STS and Non-STS sections for all fifteen teacher participants who had been partners in professional development efforts for at least ten years with the Iowa Chautauqua and SS&C projects.. They were leaders in terms of work on a variety of Action Research projects and well known in other districts where the STS approach was used. They typically used STS approaches in their own classrooms. All were considered leaders for science teaching in grades 5 through 9 for the Iowa SS&C project. School administrators and counselors were positive and helpful with the research, especially the ones involved with this study. The students involved were typical for both sections for each of the teachers. School counselors reported finding no differences in terms of gender, socio-economic factors, and ability levels. One concern was related to the teaching in Non-STS sections which was not the typical style for the 15 teachers; they did try hard not to make students in the Non-STS classrooms to feel disadvantaged. Nonetheless, this could be a factor that was not standard nor observable even after analysis of classroom observations and/or via video-tapes. School counselors did report that there were no complaints from students nor parents for student not experiencing the STS approach.

The tables show clearly that there were no differences in terms of pre-test scores in all five domains in terms of the STS and Non-STS students for the fifteen teachers. On the other hand, the post test scores illustrate mostly positive changes in all five domains for students enrolled in the STS sections for each of the fifteen teachers.

As indicated in Table I student growth in terms of Concept Mastery with the post-tests was not different for the students enrolled in the two sections for each teacher. This is important since the Non-STS students focused largely on Concept Mastery while the students in the STS sections learned concepts that were needed as they worked on problems with personal and local concerns. Some teachers generally are often concerned that students will learn fewer concepts since they are not the driving forces for the lessons or unit studies. The lack of any differences is encouraging and a positive result showing that STS does not limit Concept Mastery – just because they are not used as instructional organizers.

Table I
Summary of the ANCOVA for Comparisons of Student Performances in Fifteen STS and Non-STS Classrooms Concerning the Concept Domain

Teacher	Group	n	Mean Pre	Mean Post	S.D. Pre	S.D. Post	t	p	F	p
1	STS	21	8.36	16.72	2.76	3.53	23.52	0.00	341.98	0.000
	Non-STS	19	8.69	17.69	2.61	4.79	15.33	0.00		
2	STS	14	3.03	6.55	1.37	1.94	22.68	0.00	407.70	0.000
	Non-STS	17	2.57	6.80	1.10	1.69	16.81	0.00		
3	STS	21	6.00	13.00	2.35	3.44	22.35	0.00	1241.39	0.000
	Non-STS	24	5.70	12.40	2.20	3.21	22.33	0.00		
4	STS	29	1.68	6.59	0.89	1.84	17.15	0.00	135.58	0.000
	Non-STS	32	1.56	6.30	0.84	2.03	15.31	0.00		
5	STS	16	2.42	6.50	1.41	2.38	15.35	0.00	391.64	0.000
	Non-STS	14	2.24	6.60	1.50	2.56	14.56	0.00		
6	STS	26	2.85	9.09	1.52	3.01	14.88	0.00	245.68	0.000
	Non-STS	21	2.95	9.54	1.25	2.66	17.80	0.00		
7	STS	16	2.42	7.46	1.41	2.40	18.74	0.00	287.26	0.000
	Non-STS	18	2.66	6.79	1.49	2.76	11.07	0.00		
8	STS	28	4.18	11.81	2.01	3.36	13.35	0.00	186.45	0.000
	Non-STS	27	4.20	13.37	2.32	2.76	31.98	0.00		
9	STS	15	5.30	12.87	2.49	3.20	20.98	0.00	324.73	0.000
	Non-STS	16	5.54	12.83	2.68	3.47	16.88	0.00		
10	STS	23	6.48	13.29	2.35	3.36	24.55	0.00	990.08	0.000
	Non-STS	21	6.34	12.69	2.41	3.71	19.08	0.00		
11	STS	18	3.88	12.52	1.65	2.64	23.80	0.00	200.92	0.000
	Non-STS	19	4.25	12.62	1.57	2.21	23.78	0.00		
12	STS	22	4.66	13.51	2.10	3.13	28.69	0.00	571.34	0.000
	Non-STS	20	4.67	14.60	2.10	2.42	45.58	0.00		
13	STS	23	5.96	17.12	2.83	3.57	23.15	0.00	271.24	0.000
	Non-STS	21	5.29	15.51	2.09	3.90	23.00	0.00		
14	STS	21	6.04	12.68	2.97	3.79	21.14	0.00	722.36	0.000
	Non-STS	19	5.96	12.65	2.64	3.69	18.66	0.00		
15	STS	17	9.00	16.22	3.78	3.75	27.43	0.00	1424.38	0.000
	Non-STS	14	9.96	15.96	3.58	3.70	22.67	0.00		

Table II reports on the results focusing on the learning of general Process Skills. A focus on such mastery is not often a primary goal, especially at the middle school or early high school levels. Of particular importance is the fact that student learning of process skills is enhanced in the STS sections. Apparently thinking and analogies, (as well as experience with the specific fourteen processes basic to the Science-A Process Approach (SAPA) (AAAS, 1965) are realized to a greater extent in STS sections over those in Non-STS classrooms. SAPA was a K-8 program for pre-K through grade 8 classrooms in the late 60s. Significant increases for students in the STS sections could be caused by the fact that Concept Mastery is the major focus of traditional teaching and in the Non-STS classrooms of the STS teacher leaders. Typical teacher and textbook examinations focus on Process Skill Mastery per se. Similarly, typical laboratories do not focus on processes as foci for learning. Hence the results indicated in Table II are not unexpected.

Table II
Summary of the ANCOVA for Comparisons of Student Performances in Fifteen STS and Non-STS Classrooms Concerning the Process Domain

Teacher	Group		Mean		S.D.		t	p	F	p
		n	Pre	Post	Pre	Post				
1	STS	21	4.36	9.28	1.18	2.45	15.84	0.000	276.00	0.000
	Non-STS	19	4.19	4.34	1.35	1.54	0.84	0.404		
2	STS	14	2.18	5.40	0.96	1.86	13.38	0.000	293.24	0.000
	Non-STS	17	1.84	2.80	0.92	1.23	8.18	0.000		
3	STS	21	4.22	7.66	1.70	1.78	18.64	0.000	715.37	0.000
	Non-STS	24	3.50	4.10	1.60	1.83	3.94	0.000		
4	STS	29	2.09	5.72	1.06	1.83	15.60	0.000	297.08	0.000
	Non-STS	32	1.69	2.87	0.82	1.32	7.24	0.000		
5	STS	16	2.42	7.38	1.03	2.33	17.38	0.000	305.10	0.000
	Non-STS	14	2.20	3.08	1.32	1.25	6.60	0.000		
6	STS	26	2.57	8.61	1.24	2.90	14.51	0.000	308.06	0.000
	Non-STS	21	2.59	3.40	1.05	1.62	5.23	0.000		
7	STS	16	2.42	7.76	1.36	2.59	13.24	0.000	134.73	0.000
	Non-STS	18	2.62	3.95	1.24	2.07	4.87	0.000		
8	STS	28	2.90	9.09	1.37	2.36	21.20	0.000	358.26	0.000
	Non-STS	27	2.79	4.04	1.17	2.01	5.31	0.000		
9	STS	15	6.52	11.82	2.60	3.18	16.10	0.000	618.68	0.000
	Non-STS	16	6.75	7.66	2.78	3.19	3.25	0.004		
10	STS	23	3.14	9.77	1.61	2.48	27.00	0.000	729.10	0.000
	Non-STS	21	3.42	5.75	1.81	2.39	6.32	0.000		
11	STS	18	5.11	10.64	1.76	2.52	8.05	0.000	352.37	0.000
	Non-STS	19	4.81	5.68	1.90	2.30	4.34	0.001		
12	STS	22	2.72	10.03	1.30	2.59	21.23	0.000	290.63	0.000
	Non-STS	20	3.00	3.92	1.24	1.69	41.00	0.000		
13	STS	23	3.40	4.77	1.55	2.15	21.04	0.000	373.29	0.000
	Non-STS	21	3.40	4.77	1.55	2.15	7.36	0.000		
14	STS	21	5.08	9.12	2.13	2.81	12.67	0.000	553.62	0.000
	Non-STS	19	4.50	5.76	1.98	2.38	7.04	0.000		
15	STS	17	4.88	10.22	2.02	2.34	26.70	0.000	956.08	0.000
	Non-STS	14	4.76	5.92	2.14	2.44	6.45	0.000		

Table III reports on the differences in terms of student outcomes in the Application/Connection Domain. The results again clearly indicate the superiority of the STS approach in terms of applying concerns (and process skills) in new situations. Not surprisingly, the students in the Non-STS classes do not excel in applying and/or connecting their learning to anything else in their lives. They are not expected to do more than taking notes, remembering, and repeating what they are told or what they have read. One of the key advantages of the STS approach is the ability to apply information and skills in new situations. To many this is the ultimate proof of learning and something that every teacher (and pupil) should accomplish. And yet, it rarely occurs – even in classrooms where experienced and enthused STS teachers elect not to focus on any applications which are required and central to the STS approach. STS starts with problems – often related to the environment or energy needs. Non-STS situations are usually devoid of issues, problems, applications or actions in school or the lives of students outside of school.

Table III
Summary of the ANCOVA for Comparisons of Student Performances in Fifteen STS and Non-STS Classrooms Concerning the Applications Domain

Teacher	Group		Mean		S.D.		t	p	F	p
		n	Pre	Post	Pre	Post				
1	STS	21	6.32	18.32	2.01	3.80	22.54	0.000	383.03	0.000
	Non-STS	19	6.46	8.03	2.42	3.37	6.49	0.000		
2	STS	14	1.37	7.29	1.37	7.29	18.27	0.000	155.74	0.000
	Non-STS	17	1.19	2.23	0.74	1.33	8.18	0.000		
3	STS	21	5.33	13.00	2.30	3.71	19.35	0.000	632.33	0.000
	Non-STS	24	4.30	6.00	1.62	2.47	6.03	0.000		
4	STS	29	1.31	6.27	0.94	1.51	25.86	0.000	243.50	0.000
	Non-STS	32	1.34	3.52	0.88	1.44	11.13	0.000		
5	STS	16	0.76	6.00	0.71	2.24	14.16	0.000	94.76	0.000
	Non-STS	14	1.04	2.12	0.20	0.21	8.43	0.000		
6	STS	26	1.90	8.38	1.33	2.69	17.85	0.000	294.23	0.000
	Non-STS	21	1.68	2.50	1.32	1.53	5.23	0.000		
7	STS	16	1.92	6.84	1.44	2.14	18.53	0.000	278.08	0.000
	Non-STS	18	1.54	2.16	1.10	1.16	4.30	0.000		
8	STS	28	1.81	10.63	1.25	3.87	11.95	0.000	53.34	0.000
	Non-STS	27	1.95	2.54	0.99	1.50	3.44	0.002		
9	STS	15	2.21	12.04	1.78	3.02	20.58	0.000	116.07	0.000
	Non-STS	16	1.66	5.79	1.09	2.46	10.93	0.000		
10	STS	23	2.55	10.18	1.15	2.60	24.55	0.000	181.40	0.000
	Non-STS	21	2.53	4.84	1.27	2.24	7.50	0.000		
11	STS	18	2.41	11.41	1.12	3.00	16.80	0.000	101.23	0.000
	Non-STS	19	2.31	3.56	1.07	1.59	4.69	0.000		
12	STS	22	2.13	12.31	1.62	3.21	23.34	0.000	192.09	0.000
	Non-STS	20	1.78	3.00	1.10	1.36	8.70	0.000		
13	STS	23	3.12	15.12	1.61	3.19	25.02	0.000	214.39	0.000
	Non-STS	21	3.25	5.70	1.48	2.63	8.76	0.000		
14	STS	21	3.40	13.72	1.65	2.71	29.49	0.000	396.46	0.000
	Non-STS	19	3.57	6.19	1.96	2.85	9.81	0.000		
15	STS	17	4.74	15.92	2.04	3.38	31.57	0.000	643.71	0.000
	Non-STS	14	4.64	6.32	1.84	2.73	7.11	0.000		

Table IV indicates the results regarding the Creativity Domain. Although most recognize the importance of creativity and its role in scientific pursuits, it is not a facet for assessing student performance; it typically has little or no role in typical classrooms – again where Concept Mastery is the major focus. In all fifteen STS sections enhanced creativity was found to be significantly better. Students experiencing the STS approach asked more questions, raised more unique questions, offered more ideas about possible explanations for the objects and event studies – and more often they offered more unique explanations. The students in the STS sections also were able to provide evidence of the validity of some of their explanations; they were also quick to discuss their evidence with others, and to enter into debates and arguments. The STS students were also better able to suggest consequences for some predictions. In every case the STS students were better in all aspects of creativity than students in Non-STS sections where they were the receivers of information from teachers or textbooks – merely to be remembered and/or replicated.

Table IV

Summary of the ANCOVA for Comparisons of Student Performances in Fifteen STS and Non-STS Classrooms Concerning the Creativity Domain

Teacher	Group		Mean		S.D.		t	p	F	p
		n	Pre	Post	Pre	Post				
1	STS	21	23.04	52.72	8.75	16.96	15.29	0.000	450.46	0.000
	Non-STS	19	23.11	23.96	6.43	5.76	2.14	0.042		
2	STS	14	87.66	163.96	33.37	50.98	18.06	0.000	1326.39	0.000
	Non-STS	17	83.96	79.42	33.01	33.93	3.36	0.002		
3	STS	21	76.61	163.33	25.97	39.68	23.07	0.000	843.45	0.000
	Non-STS	24	76.00	74.90	23.71	20.06	0.87	0.391		
4	STS	29	54.90	115.59	20.94	36.65	17.56	0.000	1261.70	0.000
	Non-STS	32	55.65	63.43	22.07	22.40	9.05	0.000		
5	STS	16	68.42	135.76	26.20	47.66	15.33	0.000	1243.57	0.000
	Non-STS	14	65.72	72.00	28.44	32.96	15.35	0.000		
6	STS	26	76.71	133.95	25.58	34.07	20.04	0.000	1113.20	0.000
	Non-STS	21	68.50	75.59	25.70	29.28	3.52	0.002		
7	STS	16	25.30	46.92	9.78	15.59	17.60	0.000	987.44	0.000
	Non-STS	18	24.37	25.04	9.09	10.08	0.69	0.495		
8	STS	28	60.18	106.77	23.34	27.76	21.78	0.000	1669.32	0.000
	Non-STS	27	61.41	65.91	22.90	25.76	3.37	0.003		
9	STS	15	68.04	115.08	21.05	31.35	15.79	0.000	940.57	0.000
	Non-STS	16	67.37	68.20	21.18	22.64	0.67	0.509		
10	STS	23	70.33	127.96	25.31	38.78	19.36	0.000	431.21	0.000
	Non-STS	21	65.00	69.57	26.65	33.28	2.61	0.015		
11	STS	18	76.00	125.70	24.37	36.14	15.76	0.000	757.97	0.000
	Non-STS	19	71.68	72.75	24.31	23.98	0.43	0.672		
12	STS	22	24.00	41.48	8.99	12.83	20.52	0.000	1287.70	0.000
	Non-STS	20	24.39	24.42	8.16	8.82	0.05	0.962		
13	STS	23	72.36	125.76	24.11	37.23	17.56	0.000	1429.06	0.000
	Non-STS	21	69.63	74.59	26.16	29.18	2.92	0.007		
14	STS	21	66.12	113.04	21.03	27.54	24.15	0.000	1335.92	0.000
	Non-STS	19	61.57	69.23	22.35	27.61	3.76	0.001		
15	STS	17	64.92	117.14	22.57	31.88	22.83	0.000	1463.69	0.000
	Non-STS	14	63.24	62.96	23.15	23.28	0.20	0.839		

Table V is a report of the differences between students in the STS and Non-STS sections in terms of their attitudes about science study, science classes, science teachers, science concerns, and their views of science versus technology (the "natural" world vs the "design" world). Once again there is a clear difference in terms of positive attitudes concerning the whole experience with STS teaching versus a more traditional focus on Concept Mastery. Too often teachers remain unconcerned about the negative reactions of most students K-16 concerning their attitudes toward/about science study. In fact, attitudes traditionally become more negative as students progress across the years of their schooling (Yager, Akcay, Choi, Yager, 2009). The STS approach engages students in doing science and technology and depends on their ideas, their actions, their questions – as well as the interactions among their peers in a given classroom. In the Non-STS sections there was less collegiately and more of a competitive atmosphere.

Table V
Summary of the ANCOVA for Comparisons of Student Performances in Fifteen STS and Non-STS Classrooms Concerning the Attitude Domain

Teacher	Group		Mean		S.D.		t	p	F	p
		n	Pre	Post	Pre	Post				
1	STS	21	8.36	15.08	3.08	4.49	12.23	0.000	437.27	0.000
	Non-STS	19	8.69	9.07	3.27	3.74	1.13	0.266		
2	STS	14	16.77	24.77	5.32	4.40	12.32	0.000	714.67	0.000
	Non-STS	17	17.84	17.34	5.19	5.36	1.42	0.168		
3	STS	21	10.05	15.38	2.30	2.76	12.25	0.000	261.94	0.000
	Non-STS	24	10.60	10.50	2.11	2.23	0.38	0.705		
4	STS	29	9.72	15.18	2.93	2.78	11.64	0.000	281.68	0.000
	Non-STS	32	10.47	9.95	2.50	2.73	1.47	0.156		
5	STS	16	11.80	20.42	4.60	4.87	13.36	0.000	526.43	0.000
	Non-STS	14	13.96	13.76	4.95	4.78	0.48	0.635		
6	STS	26	12.47	21.09	4.19	4.71	10.32	0.000	241.50	0.000
	Non-STS	21	14.18	13.36	3.72	3.83	2.00	0.059		
7	STS	16	14.15	20.73	3.90	4.15	10.27	0.000	272.11	0.000
	Non-STS	18	14.50	14.12	3.41	4.84	0.711	0.484		
8	STS	28	14.50	21.81	4.61	4.83	11.63	0.000	490.73	0.000
	Non-STS	27	14.58	13.95	4.14	4.39	1.53	0.139		
9	STS	15	13.04	21.60	3.90	4.74	9.38	0.000	133.17	0.000
	Non-STS	16	12.25	13.25	2.55	3.61	2.26	0.034		
10	STS	23	14.63	21.55	4.17	3.96	15.09	0.000	438.52	0.000
	Non-STS	21	15.11	14.38	3.93	4.85	1.30	0.203		
11	STS	18	14.52	20.05	3.62	3.19	8.05	0.000	256.85	0.000
	Non-STS	19	14.25	14.81	3.33	3.58	1.59	0.132		
12	STS	22	14.44	19.58	3.68	3.54	12.96	0.000	465.80	0.000
	Non-STS	20	14.92	14.35	3.18	4.02	1.27	0.212		
13	STS	23	14.16	20.00	3.98	4.07	7.36	0.000	276.49	0.000
	Non-STS	21	15.55	14.37	4.30	4.36	2.63	0.014		
14	STS	21	14.12	22.04	3.77	3.43	15.15	0.000	334.96	0.000
	Non-STS	19	15.00	13.61	3.82	3.92	2.54	0.017		
15	STS	17	15.29	20.70	3.40	3.11	15.05	0.000	687.04	0.000
	Non-STS	14	14.56	14.48	3.21	4.04	0.25	0.802		

Discussion

A Look at the Results and the Meaning They Suggest for Teaching Science

Too often achievement in science is based on the conceptual information students seem to possess as measured by standard instruments or those provided in teacher editions of standard textbooks. Although inquiry is often espoused, it is rarely tested as a form of content and/or used to indicate learning per se. Also, it is important to note the four levels of inquiry in the NRC, 2000, (p. 29) monograph. It is remarkable that many science educators maintain that open inquiry cannot be approached even in college classrooms. However, it occurred in all STS sections of the teachers in this study. The results of the study using results from assessments and learning in five of the six domains indicate considerable advantage for STS (as defined by NSTA) in all domains except Concept Mastery. As indicated no data are reported in this study concerning the 6^{th} Domain (Worldview) because of the differences in grade levels and varying research protocols to collect such information. In terms of concepts, however, there is no advantage over direct teaching and the added time that could be spent on Concept Mastery which probably resulted in more time practicing definitions and concepts directly by teacher actions/lectures and/or textbook reliance and teacher directed laboratories and demonstrations.

The results indicate that there are significant advantages for STS teaching while uncovering no disadvantages. More teachers need the assurance that nothing is lost and that it is actually easier and more fun to involve students more in planning lessons, selecting projects, identifying topics and problems to pursue. It is also possible to actually use science projects for improving and/or resolving problems identified in schools and the local communities. STS provides pathways for the use of concepts and skills instead of merely promising that they will be useful at some point in the future.

STS seems to provide a way for students to remain curious – something they have had prior to attending school as well as to having fun and working on problems they identify and about which they are concerned. It is too bad that parents, administrators, state agencies, and many funding groups continue to argue about identifying concepts and needed skills (often merely focusing on terms claimed to be needed for future endeavors). These are seen as boring practices before working on real problems that are local, personally relevant, and of current importance. The starting point for science as it is for scientists themselves is not a new vocabulary or a listing of Key Concepts to be learned without a real context on any past or current student related experiences.

Conclusions

The results of the semester long teaching, which is the focus of this study, permit the following conclusions: 1) Students learn as many (occasionally more) basic concepts when approached via an STS pathway. 2) Students learn more and more useful science process skills with STS approaches than occur generally in more traditional classrooms and hence one might expect it to favor the students in Non-STS classrooms. 3) Students become even more creative as they study in an STS mode. Creativity can be defined as asking questions about the objects and events in the natural world as opposed to "going through" a textbook or required curriculum. 4) Students develop more positive attitudes about science the more they study science with an STS approach. (In typical classrooms attitudes become more negative the longer science is studied in schools!). 5) Students learn better how to apply science concepts and skills in new contexts than when science is experienced with an STS teaching approach.

When considering the broadened view of science content as outlined in the National Standards, the STS approach is easier to use while also illustrating the visions for the reforms of teaching which are central to the standards. Once again it is apparent from the results of this study that how teachers teach is more important than what they teach. Perhaps there is still too much focus in too many schools on the "What"!

References

Abell, S. K. & Lederman, N. G. (eds.) (2007) *Handbook of research on science education*. Mahwah, NJ: Lawrence Erlbaum Associates.

American Association for the Advancement of Science (AAAS). (1990). *Science for all Americans.* Washington, D.C.: Author.

American Association for the Advancement of Science (AAAS, 1965). The psychological bases of Science—A Process Approach (SAPA). Washington, DC.

Anderson, R. D. & Mitchener, C. P. (1994). Research on science teacher education. In D. L. Gabel (Ed.), *Handbook on Science Teaching and Learning* (pp. 3-44). New York: MacMillan.

Bybee, R. W. (1991). Science-technology-society in science curriculum: The policy practice gap. *Theory into Practice,* 30(4), 294-302.

Bybee, R. W. & Bonnstetter, R. J. (1985). STS: What do the teachers think? In R. W. Bybee (Ed.), *NSTA Yearbook: Science/technology/society (pp. 117-127). Washington, DC: National Science Teachers Association.*

Clark, C. M. & Peterson, P. (1986). Teachers' thought processes. In M. C.Wittrock (Ed.). *Hand book of research on teaching* (3rd. ed., pp. 255-296). New York: Macmillan.

Enger, S. K. (1997). *The relationship between science learning opportunities and Ninth grade students' performance on a set of open-ended science questions.U*npublished doctoral dissertation, The University of Iowa, Iowa City.

Enger, S. K. & Yager, R. E. (2009). *Assessing student understanding in science: A Standards-based K-12 handbook (Second Edition).* Corwin Press, Inc.,Thousand Oaks, CA

Enger, S. K. & Yager, R. E. (2001). *Assessing student understanding in science: A standards-based K-12 handbook.* Corwin Press, Inc. Thousand Oaks, CA

Good,T.L. & Brophy, J.E. (2000). Looking in classrooms. (8th ed.). New York, NY: Addison-Wesley Educational Publishers.

Harms, N. C. (1977). Project Synthesis: An interpretative consolidation of research identifying needs in natural science education. (A proposal prepared for the National Science Foundation.) Boulder: University of Colorado.

Harms, N. C. & Yager, R. E. (eds.) (1981). *What research says to the science teacher, Vol. 3.* Washington, D.C.: National Science Teachers Association.

Kimble, L. (1999). *A comparison of observed teaching practices with teacher perceptions of their teaching during and following major funding.* Unpublished Doctoral dissertation. The University of Iowa, Iowa City.

Levitt, K. E. (2002). An analysis of elementary teachers' beliefs regarding the teaching and learning of science. *Science Education,* 86(1): 1-22.

Luft, J. A. (2001). Changing inquiry practice and beliefs? The impact of a one-year inquiry-based professional development program on secondary science teachers. *International Journal of Science Education,* 23, 517-534.

Luft, J. A., Roehrig, G. H., & Patterson, N. C. (2003). Contrasting landscapes: A comparison of the impact of different induction programs on beginning secondary science teachers' practices, beliefs, and experiences. *Journal of Research in Science Teaching, 40, 77-97.*

Massenzio, L (2001). Teacher beliefs about teaching science through science-technology-society. Unpublished doctoral dissertation. Georgia State University, Atlanta, GA.

Mitchener, C. P. & Anderson, R. (1989). Teachers' perspective: Developing and implementing an STS curriculum. *Journal of Research in Science Teaching,* 26(4), 351-369.

Mitman, A. L., Mergendoller, J. R., Marchman, V. A. & Packer, M. J. (1987). Instruction addressing the components of scientific literacy and its relation to student outcomes. A*merican Educational Research Journal,* 24(4), 611-633.

National Assessment of Educational Progress. (1978). *The third assessment of science*, 1976-77.

National Research Council (NRS) (2000). *Inquiry and the National Science Education Standards: A Guide for Teaching and Learning.* Washington, DC: National Academy Press.

National Research Council (NRC). (1996). *National Science Education Standards: observe, interact, change, learn.* Washington, DC: National Academy Press.

National Science Teachers Association (NSTA) (2008-09). Science/technology/ society: A new effort for providing appropriate science for all. The NSTA position statement, pp. 242-244. NSTA Handbook, Arlington, VA

National Science Teachers Association (NSTA) (1990). Science/technology/society: A new effort for providing appropriate science for all (The NSTA position statement). *Bulletin of Science, Technology and Society*, 10(5&6), 249-250.

Program Effectiveness Panel (PEP). (1992). The Iowa Chautauqua Program: A model for effecting change in the teaching and learning of science in schools. Proposal to the Department of Education, Washington, DC.

Rosenshine, B. (1971). *Teaching behaviors and student achievement.* London: National Foundation for Educational Research.

Rubba, P. A. (1989). An investigation of the semantic meaning assigned to concepts affiliated with STS education and of STS instructional practices among a sample of exemplary science teachers. *Journal of Research in Science Teaching,* 26(8), 687-702.

Tillotson, J. (2005). Investigating the meaningfulness of pre-service programs across the continuum of teaching (IMPPACT). (A proposal prepared for The National Science Foundation.) Syracuse University, Syracuse, NY.

Tillotson, J. (1996). *A study of the links between features of a science teacher preparation program and new teacher performance with regard to constructivist teaching.* Unpublished doctoral dissertation. The University of Iowa, Iowa City.

Varrella, G. (1997). *The relationships of science teachers' beliefs and practice.* Unpublished doctoral dissertation. The University of Iowa, Iowa City.

Weiss, I. R. (1993). Science teachers rely on the textbook. In R. E. Yager (Ed.). *What research says to the science teacher, volume seven: The science, technology, society movement.* Washington, D.C., National Science Teachers Association.

Weiss, I. R. (1994). A profile of science and mathematics education in the United States: 1993. Chapel Hill, NC: Horizon Research Inc. (ERIC Document Reproduction Service No. ED 382 461).

Wiggins, G. P. & McTighte, J. (1998). *Understanding by design.* Alexandria, VA: Association for Supervision and Curriculum Development.

Williams, B. (1994). Teacher-Assisted STS learning. In J. Solomon & G. Aikenhand (Ed.), STS Education: international perspectives on reform (pp. 60-67).New York, NY.: Teacher College Press.

Wilsman, M.J. (1991). Summative evaluation report. Phase 3. Secondary grades STS project. Madison, WI: Wisconsin Public Telecommunications for Education.

Yager, R. E. (1982). *What research says to the science teacher, volume 4.* Washington, DC: National Science Teachers Association.

Yager, R. E. (1993). *What research says to the science teacher, volume 7.* Washington, DC: National Science Teachers Association.

Yager, R. E., Akcay, H., Choi, A., & Yager, S. O. (2009). Student success in recognizing definitions of eight terms found in fourth grade science textbooks. *Electronic Journal of Science Education. Vol. 13, No. 2.*

Yager, R. E., Kellerman, L. & Blunck, S. M. (1992). *The Iowa Assessment Handbook.* The University of Iowa, Science Education Center, Iowa City IA.

Yager, R. E. & McCormack, A. J. (1989). Assessing teaching/learning successes in multiple domains of science and science education. *Science Education,* 73(1), 45-58.

Yager, R. E. & Penick, J. E. (1990). Science teacher education. In W. R. Houston (ed), *Handbook of research on teacher education* (pp. 657-673). New York: Macmillan.

Yager, R. E. & Tamir, P. (1992). Accomplishing STS instruction and the student learning that results. In M. O. Thirunarayanan (ed.), *Handbook of science, technology and society, Volume I: A theoretical and conceptual overview of science, technology and society education.* Tempe, Ariz.: Arizona State University, College of Education.

Yager, R. E. & Weld, J. D. (1999). Scope, sequence and coordination: The Iowa project, a national reform effort in the USA. *International Journal of Science Education,* Vol. 21, No. 2, pp. 169-194.

Yager, R. E. & Yager, S. O. (2007). What results indicate concerning the successes with STS instruction. *Science Educator,* Spring 2007, Vol. 16, No. 1.

Yeany R. H. & Padilla, M. J. (1986). Training science teachers to utilize better teaching strategies: A research synthesis. *Journal of Research in Science Teaching,* 23, 85-95.

Yutakom, N. (1997). *The congruence of perceptions and behaviors exhibited by twelve successful middle school teachers in implementing science / technology / society /*

constructivist practices in Iowa scope sequence, and coordination schools. Unpublished doctoral dissertation. The University of Iowa, Iowa City.