ISSN 1937-1831

Spring/Summer 2015
Volume 10 Number 1

# JOURNAL OF VIRGINIA SCIENCE EDUCATION

**Journal of Virginia Science Education**
The peer-reviewed journal of the Virginia Association of Science Teachers

Spring/Summer 2015 Volume 10 Number 1

## Theme: Collaboration, Professional Development & Inquiry in the Age of Standardization

**Journal of Virginia Science Education**
The peer-reviewed journal of the Virginia Association of Science Teachers

Spring/Summer 2015 Volume 10 Number 1

**Special Addendum: Vexations & Ventures, Virginia Science Education at the Crossroads: Standardized Testing**



**Review of Current Publications Affecting the Teaching of Science**

**Journal of Virginia Science Education**
The peer-reviewed journal of the Virginia Association of Science Teachers

Spring/Summer 2015 Volume 10 Number 1



## From the Editor

Collaboration is an important part of expanding one's perspective on education. As I meet with educators, including teachers, specialists, principals, district level administrators and superintendents, I realize that everyone has the same goal: to increase academic achievement. Where the differences lie is that each individual has a different manner in which they seek to achieve this outcome. For instance, the science teachers would like administrators to understand that they just want to teach science. The science specialists want to make sure that science content is being taught in alignment with state standards. The principals and administrators strive to attain their schools' accreditation and to achieve excellence. Thus, the question is: can all groups succeed as more demands are made on the outcomes?

To gather more information about how these standards and testing can affect certain groups, I asked several educators about their opinions and experiences with testing. Their results are summarized in the last section of this journal called: Vexations and Ventures, Virginia Science Education at the Crossroads: Standardized Testing. In this section, the educators voice their concerns (Vexation) and their path to better understand the concern or change it (Venture). It is my hope that you find this addendum to the regular journal interesting and that it will spark basic conversations with other educators. There is an annual national academic conference you can attend where you can write and present a Vexation and Venture paper called Science Education at the Crossroads. It is currently funded by the National Science Foundation, designed to enable various stakeholders in science education to confer in a legitimate and productive manner. The conference also claims to offer an alternative to the standard model of conferring and professional development. Conference creators and coordinators are John Settlage of the University of Connecticut and Adam Johnston of Weber State University.

On another note, some of the papers in this journal discuss how science education is approached from as early as the pre-school environment up to the college level. As always, we are asking you to think about submitting to the journal. Read our upcoming themes and consider contributing to the science community. If you have other articles that are not based on our themes, we will also consider those on an on-going basis.

Cheers,
Anne Mannarino

**Journal of Virginia Science Education**
The peer-reviewed journal of the Virginia Association of Science Teachers

Spring/Summer 2015 Volume 10 Number 1

## Call for Submissions

### Fall 2015/Winter 2016 – Getting Students Engaged in Science Using Distance Learning and Technology

The students of today are digital natives. Technology is all around them and is part of their academic, social, and learning environments. As science educators we need to know how to relate to these digital natives and teach the science content. How can we make science relevant to students using technology and/or distance learning? What have you found that works for you? How do we maintain teaching based on strict content standards and still create an interactive environment that allows for creativity, problem solving, and learning to occur in this digital age? What is the role of the teacher in getting students interested and engaged in science using technology and/or distance learning? How do we really know that the students have made connections and understand the science concepts? In this journal, we seek articles describing how science educators successfully manage to use technology and/or distance learning to teach science to experienced digital natives.

**The deadline for submission is December 31, 2015.**

### Spring/Summer 2016– Creating a Science Classroom for All Children

Teaching science is not an easy task. The demands on teachers can be overwhelming. In addition to the daily administrative tasks, the dynamics of our classroom populations are forever changing. Students may come to us with different challenges and diverse backgrounds. As teachers we are expected to support all students and teach science. How do you achieve that in your classroom? What challenges are you experiencing and how can you deal with students that learn in different ways? Diversity encompasses race, gender, socio-economic status, levels of background knowledge and experiences, learning styles, abilities, success with education, nationality, and much more. With such diversity sometimes comes a challenge to reach all children. How can we create a classroom for all children if they have different cultural backgrounds, language challenges, learning needs, behavioral issues, and family issues? In this journal, we seek articles describing how science educators manage to overcome these issues and create a science classroom for all children.

**The deadline for submission is March 31, 2016**

Additionally, we are always accepting submissions outside the current call.

We are developing a more comprehensive list of upcoming themes for the Journal. If there are themes you wish to see highlighted, please develop a Call for Submissions and submit it for consideration to journal@vast.org.

If you have questions after reviewing the information on the web site, feel free to email Anne Mannarino at journal@vast.org.

**Journal of Virginia Science Education**
The peer-reviewed journal of the Virginia Association of Science Teachers

Spring/Summer 2015 Volume 10 Number 1

## Thank you to our reviewers

- Randy Bell
- Chris Carter
- Linda Chin
- Greg Corder
- Marilyn Elder
- Beth Folden
- Jean Foss
- Kathy Frame
- Wendy Frazier
- Debbie Hamilton
- Melissa Hamilton
- Sally Hurlbert
- Charles Jervis
- Jimmy Johnson
- Tim Johnson
- Bonnie Keller
- Suzanne Kirk
- Molly Logerwell
- Catherine Meechan
- Erin Peters
- David Slykhuis

- VAST is a comprehensive educational organization dedicated to the nurturing and advancement of science education.
- VAST is a nonprofit organization by educators for educators. It is affiliated with the Virginia Math and Science Coalition and is a State Chapter of the National Science Teachers Association.
- Consider making a tax-deductible donation to VAST to help continue to nurture and advance science education in Virginia. All donations to Virginia Association of Science Teachers, Inc. are deductible for income tax purposes. Fed ID# 54-1265890 Organization type 501c3. Contact / mail to: Jimmy Johnson, Treasurer, VAST, 12141 Winns Church Road, Glen Allen, VA 23059 treasurer@vast.org.

# Virginia Standards of Learning Data as a Means of Measuring Classroom Impact in a Teacher Professional Development Project

Turina Lewis

Tim Loboschefski

Arlene Vinion-Dubiel

Hank Yochum

James Alouf

Jill N. Granger

**Abstract**

This study examines the potential for using Virginia Standards of Learning test results as a quantitative measure of the effects of a teacher professional development project on student achievement. During the 2010-2011 academic year, thirteen elementary and middle school teachers participated in a yearlong professional development project on teaching science and math using an inquiry strategy. Teachers designed, implemented in their classrooms, and assessed four inquiry-based lessons throughout the school year. Virginia Standards of Learning test data for ten of the teachers for the years prior and after the project were obtained from collaborating school divisions in Central Virginia and analyzed for evidence of the effects of teachers' participation in the professional development project on their students' learning. Test questions that were directly related to the inquiry lessons implemented were analyzed separately to monitor potential impact of an inquiry strategy on student learning. Results of the analyses of the test data are mixed. Multiple challenges to using Virginia Standards of Learning test data to measure impact of teacher professional development on student learning are discussed. The study calls into question the validity of using aggregate standardized testing data as a meaningful measure of the quality of teacher professional development initiatives.

**Introduction**

The Standards of Learning for Virginia Public Schools (SOL) were first introduced in 1995 as a framework to raise student achievement for all students in Virginia. The subject-specific standards have been twice modified with the most recent modifications in 2009 for Mathematics and 2010 for Science (Virginia Department of Education (VDOE), 2009, 2010). One strand of the Science SOL is Scientific Investigation, Reasoning, and Logic. This strand focuses on the processes of science and encompasses skills that are necessary for engaging in scientific investigations. Likewise, the Mathematics SOL have five goals that include process skills such as Mathematical Reasoning in which an argument is analyzed to determine whether conclusions are valid and Mathematical Connections which make connections to other subject areas, like science.

An inquiry-based teaching strategy aligns with the Scientific Investigation, Reasoning, and Logic strand of the Science SOL and with the Goals found in the Mathematics SOL. Inquiry as a teaching strategy encourages students to utilize the processes of science and mathematics as they analyze data to answer a research question that is testable in a classroom environment (Virginia Mathematics and Science Coalition Task Force (VMSC), 2010). Inquiry-based teaching is highly engaging and provides students with a context for understanding the nature of science and improving math and science process skills while also learning content.

Professional development (PD) experiences for teachers that focus on inquiry-based teaching in science and math provide teachers with many opportunities for growth. Not only do these PD programs provide opportunity to explore ways of teaching, but they challenge the teachers' own understanding of the processes and nature of science and requires that the teacher acknowledge and address critical gaps in content knowledge. High quality PD programs include significant follow-up during the academic year in the classroom and require that the teachers practice and integrate new approaches to teaching with students in their classrooms. Practice is the key to inquiry teaching, because once the teacher understands inquiry as a way of thinking about knowledge acquisition, they will be able to teach students to do the same.

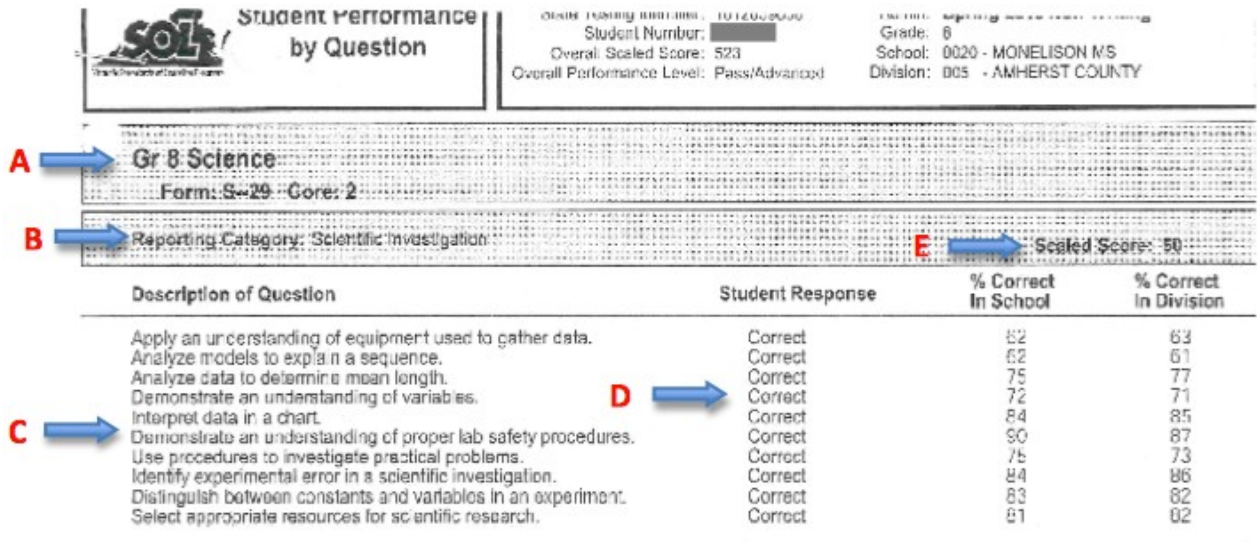
Despite positive effects on teachers' knowledge and skills, which can be readily measured through PD and evaluation plans that focus on the teacher, it is less clear how those positive effects can be measured in terms of impacts on the students of those teachers. Yet stakeholders (including federal and state funding allocators, decision-makers, administrators, and those who monitor for accountability) are very interested in the bottom-line— to what extent is students' understanding of the subject-specific content and processes outlined in the Standards. The use of state-mandated standardized testing of students has generated an enormous amount of student testing data, and also represents a huge investment and commitment on the parts of legislators to account for student learning. This study attempts to answer the question: Are the impacts of teacher professional development measurable at the student level using state-imposed standardized testing data?

While it is known that inquiry affects students' learning in positive ways (Marshall, 2014), there is very limited knowledge of how an inquiry teaching strategy affects student outcomes on large scale standardized assessments. Tretter and Jones (2004), in a small scale, multiple year study, concluded that an increase in inquiry teaching did not affect student achievement on a standardized multiple choice assessment exam. However, Geier *et.al.*

demonstrated that use of project-based inquiry science units did increase standardized test scores for middle school students in an urban setting (Geier, 2008).

The Virginia Standards of Learning tests are given towards the end of each school year to allow the students to apply their knowledge of all the material covered in the subject or course to date. Relevant to this study, SOL tests were given in science in grades 3, 5, and 8 and in math each year in grades 3-8. At the individual level, students and parents are provided with scaled score information and an indication of performance (pass, pass proficient, and pass advanced). At the school level and at the school division level, overall pass rates as a sum of all students, are available publicly for each test via the Virginia Department of Education website. These percentages are used as one of the “high stakes” measures of accountability by which schools are evaluated, and, if they fail to perform well a school may lose its accreditation (VDOE, 2014).

SOL test results are also provided to school divisions in two formats that are relevant to this study: the Student Performance By Question data (SPBQ) and the Summary Report By Question data (SRBQ). The SPBQ data is reported for each student test taker and includes a brief description of each question by reporting category and an indication if the student answered the question correctly (Figure 1). The SRBQ data summarizes the results for the school and includes the percentage of children in the school and division who answered each question correctly (Figure 2).


Student Performance by Question

Student Number: [redacted]
Overall Scaled Score: 523
Overall Performance Level: Pass/Advanced
Grade: 8
School: 0020 - MONELISON MS
Division: 005 - AMHERST COUNTY

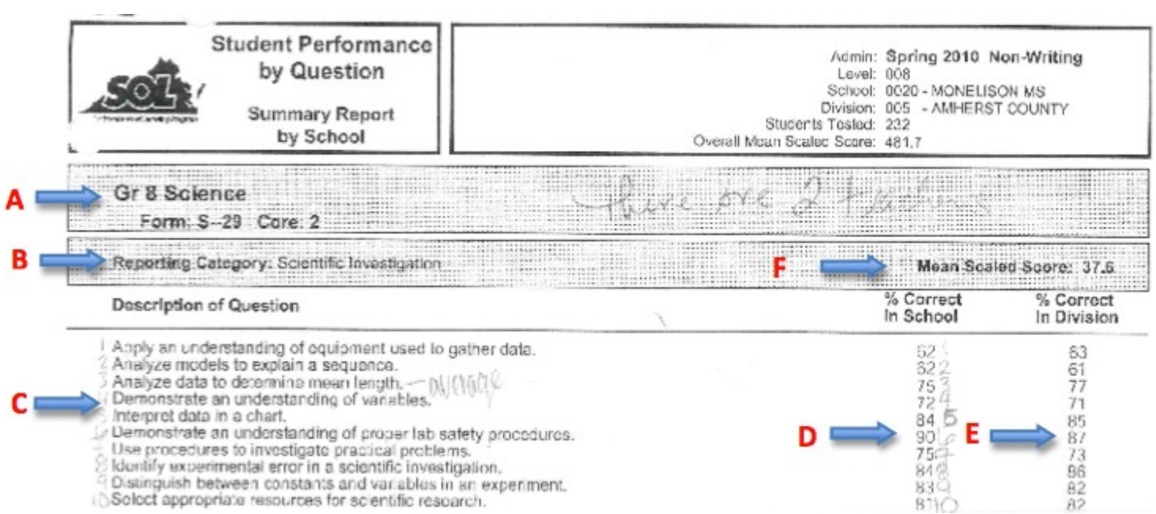
A → Gr 8 Science
Form: S-29 Core: 2

B → Reporting Category: Scientific Investigation
E → Scaled Score: 50

| Description of Question | Student Response | % Correct In School | % Correct In Division |
|---|---|---|---|
| Apply an understanding of equipment used to gather data. | Correct | 52 | 63 |
| Analyze models to explain a sequence. | Correct | 62 | 61 |
| Analyze data to determine mean length. | Correct | 75 | 77 |
| Demonstrate an understanding of variables. | Correct | 72 | 71 |
| Interpret data in a chart. | Correct | 84 | 85 |
| Demonstrate an understanding of proper lab safety procedures. | Correct | 90 | 87 |
| Use procedures to investigate practical problems. | Correct | 75 | 73 |
| Identify experimental error in a scientific investigation. | Correct | 84 | 86 |
| Distinguish between constants and variables in an experiment. | Correct | 83 | 82 |
| Select appropriate resources for scientific research. | Correct | 81 | 82 |



**Figure 1:** An example SPBQ data sheet in which the student's identifying information has been removed. In this figure, A = the grade and subject test, B = reporting category, C = description of each question in the particular reporting category, D = student's response for each question, and E = student's scaled score for this particular reporting category. In this example, items under "%Correct in School" and "%Correct in Division" are given; however, these items are not always available on each SPBQ data sheet.


Student Performance by Question
Summary Report by School

Admin: Spring 2010 Non-Writing
Level: 008
School: 0020 - MONELISON MS
Division: 005 - AMHERST COUNTY
Students Tested: 232
Overall Mean Scaled Score: 481.7

A → Gr 8 Science
Form: S-29 Core: 2

B → Reporting Category: Scientific Investigation
F → Mean Scaled Score: 37.6

| Description of Question | % Correct In School | % Correct In Division |
|---|---|---|
| Apply an understanding of equipment used to gather data. | 52 | 63 |
| Analyze models to explain a sequence. | 62 | 61 |
| Analyze data to determine mean length. | 75 | 77 |
| Demonstrate an understanding of variables. | 72 | 71 |
| Interpret data in a chart. | 84 | 85 |
| Demonstrate an understanding of proper lab safety procedures. | 90 | 87 |
| Use procedures to investigate practical problems. | 75 | 73 |
| Identify experimental error in a scientific investigation. | 84 | 86 |
| Distinguish between constants and variables in an experiment. | 83 | 82 |
| Select appropriate resources for scientific research. | 81 | 82 |



**Figure 2:** An example SRBQ data sheet. In this figure, A = the grade and subject of test, B =

reporting category, C = description of each question on the test in the particular reporting category, D = the percentage of students in the school who answered the question correctly, E = the percentage of students in the school division who answered the question correctly, and F = the mean scaled score for the school.

The purpose for this study was to use students' Virginia SOL test scores as a measure of student achievement by teacher before (2010 test event) and after (2011 test event) a year-long professional development project on inquiry instruction. Specific questions that were tested using the SOL test data included:

1. Is there an effect of teacher participation in the PD on their students' scaled scores across all reporting categories?
2. Is there an effect of teacher participation in the PD on their students' scaled scores in the category of Scientific Investigation?
3. Is there an effect of teacher participation in the PD on the number of correct answers provided by the students (rather than scaled scores) considering all the questions on the test?
4. Is there an effect on the number of correct answers provided by the students (rather than scaled scores) for the question descriptions that are most relevant to the inquiry lessons that they had experienced?
5. Is there a differential effect on the number of correct answers provided by the students as a function of relevance when comparing pre-project and post-project results?

## Methods

The PD project on which this study is based included thirteen Central Virginia teachers. Of the thirteen participating teachers, ten had suitable SOL data that could be used for this study. All ten were classroom teachers of testing grades and subjects from public school divisions in Central Virginia (Amherst, Bedford, Nelson, and Prince Edward Counties and Lynchburg City). Six of these teachers taught middle school and four elementary school. As part of the partnership with the participating school divisions, the SPBQ results for the students in the classes of the participating teachers as well as SRBQ results for the schools, for both the 2010 (pre-project comparison set) and 2011 (post-project results) testing events were provided. For the purposes of the study and to retain the anonymity of the participants, the teachers are renamed as Teacher A through Teacher J (Table 1).

Table 1: Teachers with applicable SOL testing data in the Professional Development Project

| Teacher | School Division | Grade Level | Subject |
|---|---|---|---|
| A | Prince Edward | 8th | Science |
| B | Bedford | 5th | Science |
| C | Amherst | 8th | Science |
| D | Nelson | 6th | Mathematics |
| E | Nelson | 3rd | Science |
| F | Bedford | 3rd | Mathematics |
| G | Lynchburg | 8th | Science |
| H | Amherst | 8th | Science |
| I | Lynchburg | 4th | Mathematics |
| J | Lynchburg | 8th | Science |

When data was obtained from the school divisions, it was provided either on paper or in a format that could not be modified. After discussions with school administration personnel in several of the participating school divisions, it was discovered that the information was not available in a format that could be modified electronically. Therefore, Microsoft Excel was used to digitize and organize the data prior to analysis. The following

SPBQ data (refer to Figure 1 for the categories) were transferred into electronic format: student number, scaled score, reporting category, description of question, and the correct or incorrect indicator. Correct answers were coded with a "0" and incorrect answers were coded with a "1." This data was transcribed for both 2010 and for 2011 for each teacher in the project. An example of the transcribed data is given in Figure 3.

| ***Reporting Category:*** | **Student** | | | |
|---|---|---|---|---|
| **n** | 1 | 2 | 3 | 4 |
| **Scientific Investigation:** | | | | |
| **pass/fail** | **0** | **0** | **0** | **0** |
| Q1: Understanding of equipment used to gather data | 0 | 0 | 0 | 1 |
| Q2: Analyze models to explain a sequence | 0 | 1 | 0 | 0 |
| Q3: Analyze data to determine mean length | 0 | 0 | 0 | 0 |
| Q4: Demonstrate understanding of variables | 0 | 1 | 0 | 0 |
| Q5: Interpret data in a chart | 0 | 0 | 0 | 0 |
| Q6: Understanding of proper lab safety procedures | 0 | 0 | 0 | 1 |
| Q7: Use procedures to investigate practical problems | 0 | 0 | 0 | 0 |
| Q8: Identify experimental error in a scientific investigation | 0 | 0 | 0 | 0 |
| Q9: Distinguish between constants and variables in an experiment | 0 | 0 | 0 | 0 |
| Q10: Select appropriate resources for scientific research | 0 | 0 | 0 | 1 |
| **Scale & Score** | **50** | **37** | **50** | **34** |

**Figure 3:** An example of the transcribed SPBQ data. The example shows results for the reporting category "Scientific Investigation" and shows results for students 1-4 out of 48 in Teacher A's class for the 2010 test year. The row labeled "pass/fail" indicates if the student passed (0) or failed (1) the test overall. The student's scaled score for the reporting category is shown in the bottom row.

In the PD project, the teacher-participants designed and implemented four inquiry-based lessons over the course of an academic year. Following an intense three day summer course on inquiry instruction, the teachers designed and wrote a lesson plan that used a structured inquiry approach (Bell, 2005). Instructional faculty for the PD provided feedback on the lessons which the teachers then implemented within the first four weeks of the academic year. Next, the teachers submitted reflection papers and provided assessment results. This cycle of design, feedback, implement, reflect was repeated three more times during the remaining three quarters of the academic year. The second lesson implementation followed a guided inquiry approach, the third used an open inquiry approach, and for the fourth and final inquiry, teachers were able to use any of the three approaches: structured, guided, or open (Bell, 2005). The lesson plans and reflective papers were evaluated together for their consistency with an inquiry approach using the Science Lesson Plan Analysis Instrument (SLPAI) (Jacobs, 2008). Teachers' lessons increased on average with each subsequent implementation in terms of the extent to which they were aligned with inquiry characteristics and practices as measured by the SLPAI. Thus, the aims of the PD project, in increasing these teachers' understanding of and use of inquiry pedagogy, were met (Granger, 2011).

Because the teachers implemented only four inquiry lessons over the course of the project, we used the Description of Question to align specific questions on the SOL test with the content objectives of the four inquiry lessons taught by each teacher. Each question was assigned a code: high, medium, low, or not relevant; based upon the extent to which the Description of Question was relevant to an inquiry lesson. As these were judgment calls based upon all available information, the designation for relevance was conferred by two members of the instructional faculty who had read and scored the lesson plans and reflection papers.

Microsoft Excel was used to organize the relevance data. The following information was transcribed by question for each teacher: relevance code, percentage correct for the teacher's students, percentage correct in school, percentage correct in division. This data was recorded for both 2010 and for 2011 testing events. The percentage correct for the teacher's students (Teacher's Class Average Correct) had to be calculated from the raw data, and was calculated using the expected calculation of a mean: number of correct answers divided by total number of responses, multiplied by 100. An example is given in Figure 4.

| SRBQ -Q# | Relevance | SRBQ-Reporting Category | SRBQ-Description of Question | Teacher's Class Average Correct | School's Class Average Correct | Division Average Correct: |
|---|---|---|---|---|---|---|
| Q1 | High | Scientific Investigation | Understanding of equipment used to gather data | 87.50% | 88.00% | 88% |
| Q4 | High | Scientific Investigation | Understanding of variables | 64.58% | 70.00% | 70% |
| Q5 | Medium | Scientific Investigation | Interpret data in a chart | 81.25% | 88.00% | 88% |
| Q7 | Medium | Scientific Investigation | Investigate practical problems | 77.08% | 86.00% | 86% |
| Q9 | Medium | Scientific Investigation | Distinguish b/w constants and variables in an experiment | 79.17% | 86.00% | 86% |
| Q10 | Low | Scientific Investigation | Appropriate resources for scientific research | 75.00% | 84.00% | 84% |
| Q2 | Not relevant | Scientific Investigation | Analyze models to explain a sequence | 68.75% | 70.00% | 70.00% |
| Q3 | Not relevant | Scientific Investigation | Analyze data to determine mean length | 79.17% | 84.00% | 84.00% |
| Q6 | Not relevant | Scientific Investigation | Understanding of proper lab safety procedures | 70.83% | 83.00% | 83.00% |

**Figure 4:** An example from the Excel worksheet for Teacher A in the reporting category of "Scientific Investigation." This example shows summary data for the 48 students in Teacher A's class for the 2010 test year. The Question number ("Q#"), the Reporting Category, and the Description of Question are from the SRBQ. The relevance code was a rating of the relevance of the questions to the implemented inquiry investigations as described in the text. The Teacher's Class Average Correct was calculated from the SPBQ data. The School's Class Average Correct and the Division Average Correct are from the SRBQ data.

Statistical Package Social Sciences (SPSS) was used for statistical comparisons.

Statistical comparisons were made using a combination of t-test statistics and ANOVA comparisons.

## Results

Analyses were performed for all ten teachers with applicable data. Example results for Teacher A, a grade 8 science teacher, are shown in Table 2. When the scaled scores for each student in Teacher A's class for 2010 and 2011 were averaged, a small increase in overall scaled scores was observed. When this analysis was repeated for all 10 teachers, there was no statistical gain from 2010 to 2011 across all reporting categories. ($t(9)=-0.040$, $p=0.969$)

As inquiry teaching utilizes the processes of science exemplified in the Scientific Investigation strand of the science SOLs, the mean scale scores for questions in this reporting category were compared from 2010 to 2011. The mean scores for the students of Teacher A from 2010 and 2011 increased slightly from 2010 to 2011 (Table 2). When the analysis was conducted on these scores for the entire group of teachers, there was no statistically significant gain from 2010 to 2011 in the Scientific Investigation reporting category. ($t(7)=1.182$, $p=0.282$)

The number of correct responses for the SOL tests as a percentage for the students of Teacher A in 2010 and 2011 are shown in Table 2. For Teacher A, there was an increase in the number of correct answers provided by the students overall with the questions on the tests from pre-project (2010) to post-project (2011). When this analysis was repeated for the entire group of teachers, there was no statistical gain from 2010 to 2011 when considering percent correct across all questions. ($t(11)=0.169$, $p=0.869$)

Table 2: Example SOL test data analysis of the students of Teacher A from pre-project (2010) and post-project (2011).

| | Mean Scaled Score Overall 2010 | Mean Scaled Score Overall 2011 |
|---|---|---|
| Teacher A | 37.56 | 38.48 |
| | Mean Scaled Score for Scientific Investigation 2010 | Mean Scaled Score for Scientific Investigation 2011 |
| Teacher A | 38.0 | 38.17 |
| | Percent Correct Responses Overall 2010 | Percent Correct Responses Overall 2011 |
| Teacher A | 75.96 | 79.29 |

To determine if there was an effect of inquiry teaching, specifically on the students' test results, the student's percent correct for only those questions that were identified as being relevant to the inquiry lessons taught during the project were considered. An 3 X 1 ANOVA was used to determine if there were statistical differences in the percent correct scores amongst the three categories of relevance (high, medium, and low) of the question descriptors to the content taught in the four inquiry lessons. The hypothesis was that there would be no statistical difference among categories for the pre-project data set (2010), and that any significance found among the categories for the post-project data set (2011) could indicate an effect from the inquiry lessons. Example data for Teacher A is shown in Table 3. For Teacher A, the differences among the three relevance categories for 2010 are not significant, as had been hypothesized. While 2011 was still not statistically significant ($>0.05$), the differences were much lower as seen by the p-value decrease. When this analysis was repeated for the other teachers, a wide variety of results were observed.

Teacher G showed significant difference in both pre and post data sets ($p<0.05$ for both years); defeating our hypothesis in regard to the 2010 data set. Teacher B, Teacher E, and

Teacher I showed increased p values in 2011, compared to 2010. Of interest is the fact that Teachers B, E, and I teach predominantly math. Teacher H was the only data set that was aligned with the hypothesis: 2010 p=0.606, 2011 p=0.049.

To determine if there was an effect of inquiry teaching on the student's test results, another analysis compared the average percent correct on questions designated as relevant with the questions that were designated as not relevant. The ANOVA for the students' test results from 2010 and 2011 are shown in Table 3. This analysis showed variable results for all teachers with no clear trends observed.

Table 3: Example analysis results of the significance values (p) in comparison of the percent correct for relevant questions for Teacher A's students for 2010 (pre-project) and 2011 (post-project).

| | ANOVA - Relevant Questions 2010 | ANOVA - Relevant Questions 2011 |
|---|---|---|
| Teacher A | 0.890 | 0.137 |
| | ANOVA - Relevant High compared to Not Relevant questions 2010 | ANOVA - Relevant High compared to Not Relevant questions 2011 |
| Teacher A | 0.553 | 0.848 |

**Discussion**

This study was undertaken in an attempt to answer the question: Are the impacts of teacher professional development measurable at the student level using state-imposed standardized testing data? The teachers involved in this PD demonstrated gains in knowledge and were successful at implementing inquiry-based lessons with their students (Granger, 2011). Teachers assessed student learning for each of their inquiry lessons using their own assessments. Across the board, gains in student learning were observed in these lessons. However, the quality of the assessments was not measured, so broad assumptions about the efficacy of these assessments to measure student learning must be made to make any conclusions. To standardize student learning measures, the SOL test is the most logical choice of measure to determine the impact of teacher PD on students' learning. However, a determination of the success of the PD based on students' performance on SOL tests cannot be made easily.

From an analytical perspective, this study suffered from some logistical challenges, starting with the amount of time it took to encode the data into a usable electronic format. While this study was rather small (only 10 teachers), the amount of data encoding was quite significant (10 teachers, 2 years of data, on average 30 students per teacher, typically 50 questions per test, yielded over 30,000 data entry items). One of this project's strengths was a strong collaboration between the Institution of Higher Education, Sweet Briar College, and the Local Education Agencies, the five participating school divisions. Even so, there were practical difficulties in obtaining the data. For example, project staff in some cases had to go to the school office to pull the SRBQ and SPBQ files and make copies on site. It soon became evident that the SRBQ data and SPBQ data, while they are available in the school building, are not readily provided to the teachers and are not typically familiar to office personnel. Some teachers were surprised to find out that this level of student performance results is available. This calls into question the use of this data at the division level, especially as this data is not available in a format that can be easily analyzed.

Another limitation of this study involves the logistics of linking the PD goals and activities that may influence the classroom with the test goals and questions. While overall scaled scores may increase, on average, from one year to another, the significance of that result is difficult to correlate to a particular lesson or pedagogy and is possibly due to teacher and students' familiarity with the test format. In this particular study, while positive effects on teachers' professional growth were measured directly and their understanding of, capacity for, and implementation of inquiry approaches to teaching were increased over time, the effects on students' performance on standardized tests are unclear.

Other limitations to the use of students' standardized testing data as a measure of teacher PD impact involve the exposure of the students to the influences of the PD. Teachers sometimes lose students to "pull outs" throughout the instructional period, absences, and other competing interests at the school; therefore, while the study's analysis assumes that all students received the inquiry instruction provided by the teacher, this is not rigorously true. Next, the teachers' students are not the same each year, and therefore accounting for baseline changes would be very important for a rigorous research-based approach. This study relies on the assumption that on average, the teachers' students are approximately the same in aggregate from one year to the next.

Testing discrepancies from year to year further complicate the analysis. The questions themselves on the standardized tests are not the same from one year to the next. While question descriptors and reporting categories provide some information to help make comparisons from one year to the next, the details are not available so comparisons can be imprecise. Further, due to test development at the state level, some students in the same class take different tests (with different questions) than their classmates.

Finally, quality PD is a long-term endeavor. This was the first and only year the teachers were involved in an intense PD program on inquiry instruction. Llewellyn states "In most cases, teachers may need three to five years to polish their inquiry teaching techniques." (Llewellyn, 2007, p.3). Thus, it is quite possible that the short time span of the PD program is also a likely contributing factor to the lack of clear results.

In summary, while the analytical challenges to using student SOL testing data as a measure of the effect of teacher PD, while significant, may be overcome, the use of this data by teacher over a short time span does not provide meaningful results. It is recommended that PD and measures of student impact using standardized testing be long term (minimum of 3 years). Alternatively, measures of student gains that are more specifically aligned to the goals of the PD should be used to determine student impact.

**References**

Bell, R. L., Smetana, L., & Binns, I. (2005). Simplifying inquiry instruction. *The Science Teacher*, *72*(7), 30-33.

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.

Granger, J. N., Alouf, J. L., Loboschefski, T., Vinion-Dubiel, A., Yochum, H., Herrington, D., & Yezierski, E. (2011). Assessing inquiry in STEM lessons: use of lesson plan analysis to determine the impacts of a teacher professional development program. *The Gordon Research Conference on Chemistry Education Research and Practice.* Davidson, NC.

Jacobs, C. L., Martin, S. N., & Otieno, T. C. (2008). A science lesson plan analysis instrument for formative and summative program evaluation of a teacher education program. *Science Education 92*, 1096-1126.

Llewellyn, D. (2007) *Inquire within* (2nd ed.). Thousand Oaks, California: Corwin Press.

Marshall, J. C., & Alston, D. M. (2014). Effective, sustained inquiry-based instruction promotes higher science proficiency among all groups: A 5-year analysis. *Journal of Science Teacher Education*, *25*(7), 807-821.

Tretter T. R., & Jones, M. G. (2003). Relationships between inquiry-based teaching and physical science standardized test scores. *School Science and Mathematics*, *103*(7), 345-350.

Virginia Department of Education (2014). *Accountability and Virginia public schools*. Richmond: Virginia Board of Education.

Virginia Department of Education. (2010). *Science standards of learning for Virginia public schools*. Richmond: Virginia Board of Education.

Virginia Department of Education. (2009). *Mathematics standards of learning for Virginia public schools.* Richmond: Virginia Board of Education.

Virginia Mathematics and Science Coalition Task Force. (2010). *Scientific inquiry and the nature of science.* Retrieved March 11, 2015 from http://www.vamsc.org/whitepapers.html

**Acknowledgements**

Turina Lewis thanks the Honors Program at Sweet Briar College and the Virginia Foundation for Independent Colleges (VFIC) 2012 Summer Undergraduate Science Research Fellowship Award funded by the Council of Independent Colleges for their support.

The professional development project was funded by the State Council of Higher Education for Virginia, NCLB Title II, Part A, Improving Teacher Quality State Grants, for 2010-2011.

The authors thank the following school divisions for their support in obtaining the data for this study: Amherst County Public Schools, Bedford County Public Schools, Lynchburg City Public Schools, Nelson County Public Schools, and Prince Edward County Public Schools.

# Engaging Students from Diverse Backgrounds Using Inquiry-Based Science Curriculum

Tammy R. McKeown, M.S.

Suzanne V. Kirk, M.Ed.

Lisa M. Abrams, Ph.D.

Patricia W. Slattum, Pharm.D., Ph.D.

**Abstract**

Inquiry-based instructional approaches are effective strategies to actively engaging diverse groups of learners with science content and skills. This article describes how an ongoing professional development program, Project CRESST (Enhancing Clinical Research Education for Science Students and Teachers), strives to create equitable learning opportunities for diverse populations of secondary students by providing cross-disciplinary innovative professional development experiences supported by easily integrated inquiry-based curricular materials.

## Introduction

The Next Generation Science Standards recommend using inquiry-based methods when teaching science, asserting that inquiry in science is a process requiring a range of cognitive, social, and physical activities (NGSS, 2013). Inquiry-based learning is fundamentally a question-driven approach to teaching that can be highly beneficial to students as they become actively engaged in the learning process, gain improved understanding, develop higher-order thinking skills, and acquire research skills (Spronken-Smith, Bullard, Ray, Roberts, & Keiffer, 2008). Using inquiry-based techniques in science education allows students to develop important skills such as observing and describing objects and events, formulating questions and explanations, testing hypotheses, and the ability to communicate findings. Utilizing this type of question-centered learning allows students to merge their knowledge of science with reasoning and higher-order thinking skills (National Research Council, 2006).

Although there is strong evidence that shows the benefits of inquiry-based science instruction for students from all demographic backgrounds, emerging research suggests creating real-world, student-centered knowledge may be particularly advantageous for students who identify with known academic risk factors such as low socio-economic background/poverty or English language learners (Lee, Buxton, Lewis, & LeRoy, 2005). However, teaching science in schools, especially to those in diverse urban settings that often have larger percentages of students learning English and from lower socio-economic backgrounds, can present formidable challenges. Teachers in schools serving large numbers of students from low-socioeconomic backgrounds face challenges such as overcrowding, scarce funding and resources, and a lack of high-quality science instructional materials (Lee, Buxton, Lewis, & LeRoy, 2005). The fact that students in these schools are less likely to have access to adequate instructional materials and supplies creates unequal learning opportunities. However, research has shown that students from diverse backgrounds can, in fact, learn challenging science concepts when provided with learning opportunities equal to their counterparts with more resources (Lee & Buxton, 2008).

**Overview of Project CRESST**

Project CRESST is funded by the National Institutes of Health and is designed to enhance understanding of the clinical trials process through a partnership among clinical and translational scientists, faculty in Virginia Commonwealth University (VCU) Schools of Education and Pharmacy, the VCU Center for Clinical and Translational Research, and middle and high school teachers, students, and parents. Targeted to urban and rural schools, the project features an inquiry-based curriculum for teachers to introduce students to the clinical research process, using childhood obesity research as a model. The project includes an innovative, week-long, professional development academy for secondary science and physical education teachers with follow-up throughout the school year. During the academy, teachers are engaged in activities that focus on the clinical research process and the CRESST curriculum. The project provides content-rich, inquiry-based curricular tools for teachers to introduce middle and high school life sciences students to the clinical research process within a healthy lifestyle and wellness framework.

**The CRESST Curriculum**

Although appropriate curriculum materials are essential to effective instruction, high-quality materials that meet current science education standards are scarce and are even less likely to be available in schools where diverse student populations are concentrated. There is a great need for robust science curriculum materials particularly in light of high-stakes testing and accountability in science (Lee & Buxton, 2008). The CRESST Curriculum was developed in the first year of the project and continues to be refined each year. The Curriculum Development Team for the project was comprised of experienced, highly qualified science and physical education teachers who developed classroom curricular tools, including lesson plans, exercises, and teacher notes. The curriculum is inquiry-based and aligned with state and national standards that encourage students to interact with research concepts and the clinical research process. The curriculum is divided into four main sections; (1) Research Process: General Concepts, (2) Gathering and Synthesizing Information, (3) Ethical Considerations when Conducting Research with Human Subjects, and (4) Measurement and Statistics: Collecting Data and Analyzing Results. Sample lessons include; It's All About You: A Clinical Research Simulation, What Causes Obesity and Why Do We Feel Full?, Bioethics: Standards for Scientists, and Numbers Can Talk: Measuring and Exploring Statistical Data.

**Teacher Professional Development**

In an effort to address the current need for clinical research education as well as a desire to attend to a critical health issue among adolescents, a primary aim of the project is to provide high-quality professional development to educators using an inquiry-oriented instructional approach based on content related to clinical research and health and wellness. This conceptualization of professional development – one that is content-driven, includes opportunities for active engagement and activities aligned with state and national standards -- has been shown to enhance teachers' knowledge and skills (Garet et al., 2001).

Each year Project CRESST conducts a week-long summer academy, in which middle and high school science and physical education teachers are exposed to the clinical research process through interactions with VCU faculty and investigators with ongoing research in childhood obesity, health, and wellness. Participants explore how clinical research is related to state and national science standards and how they can modify their instruction to infuse concepts of clinical research into their curriculum using inquiry-based instructional methods. Teachers' experiences in the CRESST Academy provide them with curriculum tools that can be easily integrated into their classrooms. One such activity, "A Matter of Taste," explores the relationship between genetics and food preferences. Participants are asked to rate their preference for several foods and use test strips to determine if they are genetically able to taste certain bitter compounds. Finally, participants use blue food color to dye their tongues and count the number of taste buds in a given area. By comparing the results of these three activities, they can evaluate the effects of genetics on their food preferences and diet.

As Academy participants increase their knowledge of science content and inquiry based instructional strategies, they are also provided with support and materials for implementing these new practices within their classrooms. Following the summer academy, participating teachers continue their involvement during the academic year through a follow-up independent study where they apply the CRESST curriculum and Academy experiences to their instruction, document the ways in which they modify their instruction, as well as how the lessons impact their students. Many of the teachers have noted the success of the “A Matter of Taste” lesson with their students. While the empirical data is easily observable, students often find that familial and cultural influences are as important as genetics in determining food preferences and influence dietary choices. This engages students from diverse backgrounds in discussions of nutrition, food traditions and availability within a community and opens the door to a wide variety of inquiry opportunities. Teachers are also encouraged to coordinate with other programs addressing nutrition, exercise, childhood obesity, and scientific inquiry already occurring in their schools.

**Community Engagement**

Engaging students, parents, and the community at large in discussion about clinical research and health and wellness is an important goal of the project. During the 2012 – 2013 school year, the CRESST team worked with the only full-time science teacher located in a diverse rural middle school (approximately 170 students enrolled) to help coordinate a student-led Community Health Fair. Based on the initial success of the fair, the local Health Council and other community organizations became involved and, recently, the middle school students conducted their 2$^{nd}$ annual health fair, more than doubling the number of students, booths, and participants involved. Additionally, a science/physical education team that attended the 2012 CRESST Academy from an Expeditionary Learning School worked with all of the 7$^{th}$ grade teachers to design an expedition focusing on research into healthy lifestyles. Each subject area incorporated activities and lessons designed to support the students' research of healthy lifestyle topics, (e.g., nutrition and the benefits of physical activity). The expedition culminated in a Health Fair open to students, parents, and community members.

These community health fairs used inquiry-based activities to engage student interest in health related issues in their communities. The students researched these issues and selected topics for student-led activities. They also invited community partners to provide resources to improve the health knowledge within the community. The teachers found that the topics were as varied as the students themselves and engaged learners at all levels.

**Program Assessment**

Since its inception, Project CRESST has conducted three summer Academies in which a total of 52 middle and high school teachers have participated. Of the 52 teachers, 12 were from urban schools, 23 from suburban schools, and 17 were from rural schools. As part of the project's assessment efforts, Academy participants are invited to participate in pre-post tests. Teachers complete the pre-test online before attending the Academy, and the post-test is facilitated through scheduling computer lab time for participants at the end of the week-long training.

Open ended comment sections within the post test allow participants to provide detailed feedback regarding their experience with the Academy, including elements which they found to be most valuable as well as areas for improvement. Teachers attending the Academy consistently and overwhelmingly provide positive feedback regarding the curriculum materials and professional development experience. Sample verbatim statements include: "The CRESST curriculum provides real-life and engaging activities and background information that will allow me to 'hook' students into the science concepts in our curriculum;" "I was given an opportunity to learn about the process of clinical research and inquiry-based learning. ALL of the sessions stimulated thoughts as to how I can change my lessons to make them more exciting and valuable for my students." "I now have a better understanding of how to guide my students in research. I have more tools to utilize and a better understanding of research and what the results mean and how to explain this in my personal classroom setting. I am returning to my home school with an abundance of new resources not only for my department, but for incorporating cross-curriculum into my teaching." and "I feel more confident about introducing students to the research process and how to use a real-life application to support my teaching of research."

The survey also includes a subscale comprised of 7 items that reflect teachers' confidence in teaching research concepts. Respondents selected from a five-point response scale ranging from "not Confident" to "very Confident" with mean values approaching 5 indicating greater levels of confidence. Preliminary analyses of the pre and post survey results indicate that participation in the summer professional development session has a positive impact on teachers' reported levels of confidence in their abilities to teach of research concepts and skills. Of particular relevance is teachers' increased confidence to engage students in inquiry-oriented activities. Table 1 represents responses on this subscale from 2013 Academy participants. The 2013 cohort contained the largest number of participants attending a CRESST Academy to date ($n = 23$). This cohort also marked greater inclusion of physical education teachers (43%) whereas previous Academies had been attended primarily science educators. Although responses from only one cohort are presented here, findings are indicative of previous years.

Table 1

*Change in Teachers' Confidence in Teaching Research Concepts*

| Survey Item | Pre | | | Post | | | | |
|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *n* | *M* | *SD* | *n* | *t* | *p* |
| Develop student's conceptual understanding of research | 2.93 | 1.163 | 15 | 4.20 | .862 | 15 | 5.104 | .000 |
| Have students participate in hands-on activities that meet specific research objectives | 3.33 | 1.175 | 15 | 4.53 | .516 | 15 | 4.054 | .001 |
| Engage students in inquiry-oriented activities | 3.57 | .938 | 14 | 4.50 | .650 | 14 | 3.484 | .004 |
| Engage students in applications of research in a variety of contexts | 2.88 | 1.088 | 16 | 4.13 | .719 | 16 | 5.371 | .000 |
| Facilitate students use of investigative strategies | 3.21 | .893 | 14 | 4.21 | .699 | 14 | 4.266 | .001 |
| Encourage students' interest in the research/inquiry process on a regular basis | 2.71 | .994 | 14 | 3.86 | .663 | 14 | 5.551 | .000 |
| Supervise research projects of your students | 3.20 | 1.146 | 15 | 4.00 | .926 | 15 | 3.055 | .009 |

$p < .05$

The development and implementation of high quality curricular materials is vital to the role of engaging diverse student populations in science inquiry. The benefits of inquiry-based learning have been clearly documented. Of particular importance is emerging literature, which suggests these question-centered techniques may be especially effective for students from diverse or underprivileged backgrounds. The approaches used in Project CRESST can be applied to a variety of science education classroom settings. Our project findings demonstrate the potential for inquiry-based instruction to even the playing field and engage students who may struggle in traditional education environments using non-traditional instructional techniques. What we have learned about our professional development program from participating teachers who work with diverse student populations in both urban and rural environments shows they have met with great success and their students are more engaged and enthusiastic about learning science.

***Special Thanks:***

The professional development program and research described in this article was supported by the Office of the Director of the National Institutes of Health, Science Education Partnership Award number R25 OD010983-03. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## References

Garet M., Porter A., Desimone L., Birman B., & Yoon K. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, 38, 915-946.

Lee, O., & Buxton, C. (2008). Science curriculum and student diversity: A framework for equitable learning outcomes. *The Elementary School Journal*, 109(2), 123-137.

Lee, O., Buxton, C., Lewis, S., & LeRoy, K. (2005). Science inquiry and student diversity: Enhanced abilities and continuing difficulties after and instructional intervention. *Journal of Research in Science Teaching*, 43(7), 607-636.

NGSS Lead States (2013). Next generation science standards: For states, by states. Washington, DC: The National Academies Press.

National Research Council (Ed.). (1996). *National science education standards*. National Academy Press.

Spronken-Smith, R., Bullard, J., Ray, W., Roberts, C., & Keiffer, A. (2008). Where might sand dunes be on Mars? Engaging students through inquiry –based learning in geography. *Journal of Geography in Higher Education*, 32(1), 71-86.

# Transformation of a Traditional General Biology Laboratory Course to Authentic Scientific Inquiry: Lessons Learned

Sarah M. Witiak

Tracy M. Walker

Leslie Y. Whiteman

**Abstract**

The purpose of this study was to measure the impact of scientific inquiry on students' perceptions in a redesigned general biology laboratory course. It included an evaluation of the skills and other benefits for the student, as well as the investment of time and perceived difficulty. Changes in students' perceptions were measured at the completion of the course using a pilot survey. Over a fifteen-week semester, students gradually progressed to high-level inquiry laboratory investigations with appropriate instructor scaffolding. We report the specific aspects of the course that students reported as having the greatest impact on their learning, and the elements of inquiry science that were more challenging to master. Findings demonstrated that overall, students felt positive about their experience in the course. The most strongly rated tasks related to reading scientific literature and to becoming intellectually independent. Results of the Spearman correlation found several statistically significant relationships among the students' responses.

## Background

The National Science Board estimates 96% percent of the US workforce is dependent on the 4% scientists, engineers, and entrepreneurs that drive innovations and technological developments (National Science Board, 2010). Businesses are finding it increasingly difficult to recruit qualified science, technology, engineering and mathematics (STEM) employees into their workforce (ASTD, 2009). It is in the interest of America's global competitiveness to increase the numbers of individuals with STEM degrees.

Priority recommendations to address this workforce "skills gap" crisis continue to focus on strengthening STEM education in both K-12 and post-secondary schools. Unfortunately, there is very little evidence that the US is reaching this goal. The 2012 Program for International

Student Assessment (PISA) assessment ranked US students average in science and below average in math among the world's most-developed countries (PISA, 2012). Similarly, the Trends in International Mathematics and Science Study (TIMSS) ranked US students behind many other developed nations (TIMSS, 2013). Of the recent 2013 high school graduates, only 36% of those who took the ACT reached the science benchmark (ACT, 2013).

In addition to insufficient preparation, STEM students leave the field at high rates despite the fact that "nearly 28% of high school freshmen declare interest in a STEM-related field - around 1,000,000 students each year. Of these students, over 57% will lose interest in STEM by the time they graduate from high school" (STEM Connector Report, 2012). The data from the U.S. Department of Education shows no improvement at the post-secondary level. "A total of 48 percent of bachelor's degree students and 69 percent of associate's degree students who entered STEM fields between 2003 and 2009 had left these fields by spring 2009. Roughly one-half of these leavers switched their major to a non-STEM field, and the rest of them left STEM fields by exiting college before earning a degree or certificate" (Chen, 2013). One way to improve the recruitment and retention of STEM students is by offering numerous inquiry and research experiences at all levels.

Inquiry is promoted as the gold standard for science and science education. The National Research Council (NRC) defines inquiry as "activities through which students develop knowledge and understanding of scientific ideas, as well as an understanding of how scientists study the natural world." (NRC, 1996 & 2000). In essence, science teaching should mirror the processes used by scientists. The National Science Teachers Association's (NSTA) position statement on scientific inquiry recommends that teachers help students:

- Learn how to identify and ask appropriate questions that can be answered through scientific investigations.
- Design and conduct investigations to collect the evidence needed to answer a variety of questions.
- Use appropriate equipment and tools to interpret and analyze data.
- Learn how to draw conclusions and think critically and logically to create explanations based on their evidence.
- Communicate and defend their results to their peers and others (NSTA, 2004).

The benefits of implementing inquiry in science education are numerous and well documented. Students exposed to scientific inquiry demonstrate improved scientific writing, problem-solving, questioning and critical-thinking skills (Gurwick and Krasny, 2001; Udovic, Morris, Dickman, Postlethwait & Wetherwax, 2002; DiPasquale, Mason, and Kolkhorst, 2003). Compared to traditional lab activities, students generally are more engaged and enthusiastic about inquiry activities (Ghedotti, Fielitz, and Leonard, 2004; Lord and Orkwiszewski, 2006). Inquiry-based learning precedes best in group situations that allow collaboration where knowledge is built by students in a step-wise fashion (da Cunha, 2008; Bell, Maeng & Peters, 2010).

Distinguishing authentic inquiry from an engaging hands-on activity can sometimes be difficult. The *National Science Education Standards* (NRC, 1996) evaluates scientific inquiry with two criteria questions: (1) Does the activity include a research question? and (2) Do students engage in analysis of observational or experimental data to answer the research question? The degree of inquiry that is attempted in the classroom will vary based on the abilities of the students. Students need significant scaffolding before progressing to full inquiry.

Bell, Smetana, and Binns (2005) describe inquiry-based activities as ranging from highly teacher-directed to highly student-directed, based on the amount of information provided to the student (Table 1).

Table 1. Four-level Model of Inquiry (Bell, Maeng & Peters, 2010)

| | | How much information is given to the student? | | |
|---|---|---|---|---|
| | Level of Inquiry | Question? | Methods? | Solution? |
| Teacher-Directed | 1- Confirmation | ✓✓ | ✓✓ | ✓✓ |
| ↕ | 2- Structured | ✓✓ | ✓✓ | |
| | 3- Guided | ✓✓ | | |
| Student-Directed | 4- Open | | | |

Levels 1 and 2 are more teacher-centered and generally associated with traditional "cookbook" laboratory activities. The questions and methods are given to the students, and for level 1, the expected results are already known. Levels 3 & 4 are student-centered activities that promote the learning and using of content in the process of science, and improve scientific reasoning skills. Level 2 activities can easily be converted into level 3 activities by removing the methods and having students design the experiment (Bell, Smetana, & Binns, 2005; Corder & Slykhuis, 2011). Students generate the investigative questions in level 4 open inquiry. Guided and open inquiry move students from simply remembering and understanding content to higher cognitive activities requiring the application, analysis, evaluation, and synthesis of information and ideas (Harada and Yoshina, 2004; Krathwohl, 2002).

The literature indicates students are trained to memorize textbook facts and principles without gaining any understanding of the true nature of the scientific endeavor during their high school through post-secondary education (McComas, 2007). Science curricula suffer from the lack of engaging experiences in scientific inquiry (Windschitl, 2003; Campbell & Bohn, 2008; Bell, Maeng, & Peters, 2010). The NRC called for reforming K-12 education to focus on four strands that teach the knowledge and skills for proficiency in science: 1: understanding scientific explanations; 2: generating scientific evidence; 3: reflecting on scientific knowledge; and 4: participating productively in science (Michaels, Shouse & Schweingruber, 2007). These same reforms are relevant and necessary for reforms in undergraduate science education (Handelsman, Miller & Pfund, 2007).

**Methodology: Participants and Course Description**

This study was facilitated during the fall 2013 semester at Virginia State University, a small liberal arts college in southeastern Virginia. BIOL121 Principles of Biology II is part two of a general biology course for science majors. The laboratory component of the course consists of two laboratory hours per week over a 15-week semester. All enrolled students have completed general biology part I which has a traditional laboratory component that teaches basic skills, such as using scientific equipment, measuring, observing, inferring, etc. In this study, the participants enrolled in BIOL121 laboratory course were primarily Biology or Biology Secondary Education majors and in their freshman or sophomore year of college.

Prior to this study, the original BIOL121 lab consisted of traditional cookbook (level 1 – 2) activities in which students completed preset activities. Their rigid structure had a purpose, procedure, and results section, with limited analysis, that led to one specific conclusion. The activities emphasized techniques and task completion over inductive reasoning, and promoted lower-level cognitive skills. Since the content of the general biology part II lecture course focuses on a survey of the kingdoms of life, most labs consisted of observing representative species using preserved slides or specimens. For students and faculty, these labs were generally boring and lacked relevance or authenticity.

We redesigned the course with scientific inquiry that immersed students in the nature of science, rather than following a cookbook style lab where the outcome is always predictable. The objectives of the course are to teach:

1. Use of peer-reviewed literature
2. Development of original hypotheses
3. Implementation and development of experimental protocols
4. Data analysis
5. Critical thinking
6. Scientific Writing

The redesigned course structure is quite simple – every lab session has a prelab exercise that requires the student to read the protocol before class, an inquiry activity in the classroom, and a post-lab assignment that has students analyze data that are the focus of the session (Table 2). The course is divided into three modules over the 15-week semester. Over the course of the semester, students are gradually given increasing autonomy and ownership of their lab projects to advance their **scientific** thinking and research skills.

Table 2. Biology 121 Laboratory Course Schedule

| | Pre lab assignment | Class activities | Post lab assignment |
|---|---|---|---|
| Week 1 | Introduction | Discussion of class expectations and collection of environmental bacterial samples | Hypothesis development with supportive justifications |
| Week 2 | Microscope and bacterial reproduction review | Observing bacterial plates and oral bacteria/ Gram stain | Organizing bacterial data, evaluating data for support of hypothesis and scientific drawing |
| Week 3 | Read scientific article, identify major sections | Student teams deconstruct a journal article and develop a rubric for journal evaluation | none |

| | | | |
|---|---|---|---|
| Week 4 | Background research on macroinvertebrates and nutrient effects on water quality | Hike to river with commentary on features of the river; Practice macroinvertebrate sampling | Essay on background for invertebrates, nutrients, and water quality |
| Week 5 | Sample data to practice calculations with invertebrate data | Macroinvertebrate sampling, and identification; Data entry into class spreadsheet | Written methods section for invertebrate collection and data analysis |
| Week 6 | Sample data to practice serial dilution calculations | Practice serial dilutions for phosphate test | Excel graph and caption, and calculations for unknowns |
| Week 7 | Sample data to practice calculations with phosphate data | Phosphate, pH, temperature measurements | Full draft of report |
| Week 8 | None | Peer review using class rubric | Final draft of report |
| Week 9 | Ideas for a research project | Individual students present research ideas; Research teams form and develop methods | Short PowerPoint and written proposal |
| Week 10 | None | Presentation and critique of research team's proposals | Revise proposals |
| Week 11 | None | Experiment setup and initial data collection | Draft of introduction |
| Week 12 | None | Open lab for data collection | Draft of methods |
| Week 13 | None | Open lab for data collection | Full draft of report |
| Week 14 | None | Peer review of research reports | Final report. |

During the first two-week module, students focus on prokaryotes while learning sterile technique and reinforcing microscope and observational skills by collecting bacterial samples from the campus environment. Students generate hypotheses about how campus microenvironments affect the diversity of prokaryotes. After discussing these hypotheses, students collect bacteria from an environment of particular interest to them, and then plate their samples on agar plates, which are grown at two different temperatures. The following week, students quantify and categorize the grown bacterial colonies. As every plate has a unique result, this teaches an important scientific concept that there are rarely "right" or "wrong" data sets, only more complete or more organized data sets. Students develop tables to convey the diversity of their bacterial cultures. This is a surprisingly difficult task, as the students have to make decisions about which features of the colonies are most important. The students then use their results to address their hypothesis. All of the activities require essay-type answers to give students a chance to practice and improve their writing skills.

The second module is a five-week module, where students learn about river ecology, and methods to assess water quality. Research shows that students generally enjoy field research outside of the normal lab environment (Harrill, 2000). The campus in this study is ideally located on the lower Appomattox River, a tributary of the James River and the Chesapeake Bay watershed. Faculty and students in the biology department have been involved on an ongoing Appomattox River Project to analyze water quality and evaluate its impact on a variety of organisms that live in the environment. Students monitor changes in chemical and physical properties of the water, and the subsequent effects on the local microorganisms, invertebrates, plants and animals. This type of place-based learning engages students in the local community as a focus for learning. Place-based learning has been demonstrated to help students better understand the political, economic, social, and environmental connections of their community, and emerge as better informed, active, and concerned citizens (Maguth & Hilburn, 2011).

Students initially take an introductory hike to observe the river terrain and physical features. Over the next several weeks, students collect data on invertebrates, pH, temperature, phosphate levels, and physical properties. These exercises introduce students to taxonomy, serial dilutions, and spectrophotometry. They combine this data to write a formal journal-type article. Students are led through the writing process by first examining peer-reviewed articles on water quality and then developing their own writing rubrics. The rubrics are collected and combined across all teams to create a common rubric to grade the subsequent students' reports. Working in teams, students are tasked to write an article on their Appomattox River findings that incorporate peer-reviewed references, journal-style graphs and charts, and an analysis of how the different types of data relate. They receive feedback on each section of the paper from the instructor over the several weeks of data collection, and also perform peer review of drafts from classmates using the class designed rubric.

The third module is a team project where students develop and test their own research questions. Students are told that they can use a model plant, *Lemna*, or follow up on a question that they have from the river module. The research question has to examine the influence of some environmental factor. Students are given minimal background information on *Lemna.* The first pre-lab assignment for this unit requires students to develop three potential investigative questions, to explain the motivation for the questions, and to provide a brief statement of methods that would allow them to answer the questions. During class, students chose their "favorite" question to present to their peers. After all students have presented, they are then instructed to "mix and mingle", as if at a scientific conference, and to find collaborators with similar interests. Collaborating students form research teams and design experiments to test their hypotheses. They are tasked to write a detailed formal proposal and a brief summary presentation to present during the next session. The short five-slide PowerPoint presentations have to include the background, hypothesis, methods, and expected results if the hypothesis is supported or not supported. Each session begins with the instructor's reminder of the presentation's purpose: to make sure that each team has a well-designed experiment. The class is required to ask questions and give suggestions to improve their peers' experiments. During this process, hypotheses are clarified, methods are tweaked, and controls and replicates are checked. The instructor has to approve all procedures for adherence to safety guidelines before the investigations begin. At the end of this session, students are given the opportunity to revise their research proposals. The teams are given several weeks to carry out their experiments, largely on their own, with minimal instructor intervention. During this period, research teams draft portions of their final paper and receive instructor feedback.

**Methodology: Analysis of Results**

**Instrumentation**

The current research study attempts to measure the attitudes of students enrolled in an entry-level general biology laboratory course that was redesigned to focus on inquiry exercises. This study includes an evaluation of the skills and other benefits for the student, as well as the investment of time and perceived difficulty.

The brief survey instrument was developed by the faculty teaching the BIOL121 course. The instrument consisted of five items, asking the students to rate their level of agreement according to a five-point Likert scale (1 = strongly disagree to 5 = strongly agree). The five survey items are found in Table 3 below.

Initial analysis of the survey results consisted of measures of central tendency. In particular, the mean and standard deviation were reviewed to determine the level of agreement for students as related to each of the five survey items (Table 3). The highest rated mean score was for that of 'Required me to read scientific literature' (M = 4.35, SD = 0.85). The second strongest agreement was for 'Encouraged me to be intellectual independent' (M = 4.29, SD = .970). The lowest rated item was 'Challenged me to think critically' (M = 3.38, SD = 1.23).

Interestingly, the two items in which students had the highest levels agreement also demonstrated the smallest standard deviations, while the lowest rated item had the largest standard deviation.

Table 3. Mean and Standard Deviation Results from Survey

| Survey Item | Mean | Std. Deviation |
|---|---|---|
| Developed my writing skills | 3.76 | 1.10 |
| Required me to read scientific literature | 4.35 | 0.85 |
| Increased my interest in biology | 3.24 | 1.10 |
| Encouraged me to be intellectually independent | 4.29 | 0.97 |
| Challenged me to think critically | 3.38 | 1.23 |

Note: *n* = 34

In addition to calculating the mean and standard deviation for students' ratings, the researchers also ran a correlation coefficient to determine if any relationships existed between responses. The Spearman correlation was run based on results of the tests of normality (Field, 2013). Results of the Spearman are found in Table 4. The top statistics in each row are the results of the Spearman correlation. The strongest, positive correlation is for that of 'Required me to read scientific literature' and 'Encouraged me to be intellectually independent' ($r_s$ = .556, p = .001). The second highest correlation was for that of 'Required me to read scientific literature' and 'Challenged me to think critically' ($r_s$ = .426, p = .012). The final statistically significant correlation was for that of 'Encouraged me to be intellectually independent' and 'Challenged me to think critically' ($r_s$ = ..441, p = .009). Although not statistically significant, there were moderately positive correlations demonstrated for several other items as well: 'Developed my writing skills' and 'Encouraged me to be intellectually independent' ($r_s$ = .333, p = .054); 'Increased my interest in biology' and 'Encouraged me to be intellectually independent'

($r_s$ = .332, p = .055); and 'Increased my interest in biology' and ''Encouraged me to be intellectually independent' ($r_s$ = .307, p = .078);

Table 4. Spearman Correlation Coefficient Results

| Survey Item | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Developed my writing skills | - | -.002<br>.990 | .155<br>.382 | .333<br>.054 | .157<br>.376 |
| Required me to read scientific literature | | - | .215<br>.222 | .556**<br>.001 | .426*<br>.012 |
| Increased my interest in biology | | | - | .307<br>.078 | .332<br>.055 |
| Encouraged me to be intellectually independent | | | | - | .441**<br>.009 |
| Challenged me to think critically | | | | | - |

**= Correlation is significant at the 0.01 level (2-tailed).

*= Correlation is significant at the 0.05 level (2-tailed).

**Discussion and Implications for Undergraduate Science Education**

Inquiry-based learning is based on students identifying and refining questions as they solve problems in the manner of scientist. As instructors, we would love to teach a classroom of science students who are actively engaged in the learning process, and prepared to think critically beyond the superficial knowledge and comprehension cognitive levels. We want our students to retain the content learned as they progress to upper-level science courses in their curriculum, to make connections between new information and previous ideas, and to raise questions when they sense a conflict between both. Unfortunately, in our actual classrooms, most of our students fall short of this ideal because they have not been trained in an engaging, student-centered, active-learning environment.

Students need to be able to understand and apply scientific knowledge in a meaningful way. An inability to see the relevance of classroom activities to their chosen careers (Glynn, Taasoobshirazi, & Brickham, 2009) and a lack of autonomy (Reeve and Jang, 2006; Reeve, 2009) are associated with low student motivation. The most authentic inquiry activities allow students to answer their own research questions by analyzing data they collect independently (Bell, Smetana, & Binns, 2005). At first, students are hesitant to propose truly untested hypotheses, but as they work on their ideas, they begin to see that the scientific process is about testing the hypothesis, rather than getting a “right” or “wrong” answer. By the end, students are proud of their work and begin to have the sense that they are “real” scientists. This pride may translate into enhanced motivation and courage to apply to extramural research programs and to partner with faculty in the department. From our experience, science majors that engage in authentic research are more likely to graduate with a STEM degree and pursue graduate school.

We found that gradually increasing the degree of student-directed inquiry over the course of the semester, with appropriate instructor scaffolding, prepared our students for their culminating open-inquiry research projects. Students were introduced to research using Internet resources and peer-reviewed sources. Writing skills were developed through a series of smaller assignments that allowed students to master organization, scientific style and tone. These smaller assignments with frequent feedback, also, had the potential to boost student confidence, as they saw their writing and analysis skills improve from week to week.

We found our students' attitudes about being in the lab improved with their confidence in carrying out their inquiry projects. They expressed more pleasure in performing the lab work, and seemed to grasp the relevance of the science content. Other studies have reported similar findings with regards to students' self-efficacy and inquiry (Simpson, Koballa, Oliver & Crawley, 1994; Sadeh & Zion, 2012). Anecdotally, students report that these activities challenge their writing and critical thinking skills, although we have not done formal assessments of the effect of these exercises.

Learning and utilizing scientific inquiry requires time and commitment from instructors and students. The largest challenge for the instructor in this redesigned lab format is grading the numerous writing assignments. Most of the writing exercises are fairly short, and can be done in teams to reduce this workload. Changing the pedagogy of science instruction is often problematic because instructors may lack the skills and confidence needed to manage student-directed classrooms. Frequently, adjunct or junior faculty assigned to teach entry-level biology courses are hired and placed immediately into the classroom without time for professional development. However, these challenges can be overcome through weekly lab prep meetings to discuss the objectives and strategies for grading. The benefits to student learning far outweigh the challenges from the implementation of an inquiry lab course.

**Acknowledgments**

The authors thank the students for engaging in the course and completing the surveys.

*Sarah M. Witiak, Ph.D. is an Assistant Professor in Department of Biology, College of Natural and Health Sciences at Virginia State University. Dr. Witiak is the primary instructor in the BIOL121 Principle of Biology laboratory course. Correspondence concerning this article should be sent to Sarah M. Witiak, Virginia State University, Department of Biology, P.O. Box 9064, VSU, VA 23806. Email: switiak@vsu.edu*

*Tracy M. Walker, Ph.D., is an Associate Professor in the Department of Doctoral Studies, College of Education at Virginia State University. Dr. Walker is the program evaluator for this study.*

*Leslie Y. Whiteman, Ph.D., is an Associate Professor in the Department of Biology, College of Natural and Health Sciences at Virginia State University. Dr. Whiteman is the general biology coordinator and science educator.*

## References

ACT College Readiness Report (2013): Retrieved from https://www.act.org/research/policymakers/cccr13/pdf/CCCR13-NationalReadinessRpt.pdf

American Society for Training & Development (ASTD) news report *Bridging the Skills Gap* (2009): Retrieved from http://www.astd.org/Publications/Magazines/The-Public-Manager/Archives/2009/10/Bridging-the-Skills-Gap-Part-I

Bell, R.L., Maeng, J. L. & Peters, E.E. (2010). Scientific Inquiry and the Nature of Science Task Force Report. *Virginia Mathematics and Science Coalition.*

Bell, R.L., Smetana, L, & Binns, I. (2005). Simplifying inquiry instruction: Assessing the inquiry level of classroom activities. *The Science Teacher*, Vol. *72*(7), p. 30-33.

Campbell, T., & Bohn, C. (2008). Science laboratory experiences of high school students across one state in the U.S.: Descriptive research from the classroom. *Science Educator,17,* 36–48.

Chen, X. (2013). STEM Attrition: College Students' Paths Into and Out of STEM Fields (NCES 2014-001). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Corder, G. & Slykhuis, J. (2011). Shifting to an Inquiry-Based Experience. *Science and Children*, *48*(9) 60 - 63.

da Cunha, M. V. (2008). *Chapter Five: Logic, Intelligence and Education in Dewey and Piaget.* Pragmatism, Education & Children Editions. p 97-110.

DiPasquale, D.M., Mason, C.L. & Kolkhorst, F.W. (2003). Exercise in inquiry. Journal of College Science Teaching 32 (6): 388–393.

Field, A., (2013). *Discovering statistics using IBM SPSS statistics*. Thousand Oaks, CA: Sage Publishing.

Ghedotti, M.J., Fielitz, C. & Leonard IV, D.J. (2004). Using independent research projects to foster learning in the comparative vertebrate anatomy laboratory. Bioscene 30 (1): 19–24.

Glynn, S. M., Taasoobshirazi, G. & Brickham, P. (2009) Science Motivation Questionnaire: Construct Validation with Nonscience Majors. *J. Res. Sci. Teaching*, 46 (2), 127-146

Gurwick, N.P., & Krasny, M.E. (2001). Enhancing student understanding of environmental sciences research. The American Biology Teacher. 63 (4): 236–241.

Handelsman, J., Miller, S., & Pfund, C. (2007). *Scientific teaching*. New York, NY: Freeman.

Harada, V. H., & Yoshina, J.M. (2004). Inquiry Learning through Librarian-Teacher Partnerships. Linworth.

Harrill, R. W. (2000). Evolving curricula in the new century. Journal of College Science Teaching,29, 401–407.

Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy:An Overview. *Theory Into Practice*. Vol. 41 (4), p. 212

Lord, T., & Orkwiszewski, T. (2006). Moving from didactic to inquiry-based instruction in a science laboratory. The American Biology Teacher 68 (6): 342–345.

Maguth, B. M. & Hilburn, J. (2011) *The Community As A Learning Laboratory: Using Place-Based Education to Foster A Global Perspective in the Social Studies*. Ohio Social Studies Review; 2011, Vol. 47 Issue 1, p27-34.

McComas, W. (2007). Biology Education: Under the Microscope. *Science Teacher*, *74*(7), 28-33.

Michaels, S., Shouse, A.W. & Schweingruber, H. A. (2007). Four Strands of Science Learning. *Ready, Set, SCIENCE!: Putting Research to Work in K-8 Science Classrooms*. Washington, DC: The National Academies Press.

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

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

National Science Board. *Science and Engineering Indicators.* (2010). Arlington, VA: National Science Foundation (NSB 10-01), Figure 3-3

National Science Teachers Association (NSTA) Position Statement Scientific Inquiry. (2004). Retrieved from http://www.nsta.org/about/positions/inquiry.aspx

Program for International Student Assessment (PISA). (2012): Retrieved from http://www.oecd.org/pisa/keyfindings/pisa-2012-results.htm

Reeve, J. (2009) Why Teachers adopt a Controlling Motivating Style Toward Students and How They can Become More Autonomy Supportive. *Educational Psychologist*, 44 (3), 159-175.

Reeve, J. & Jang, H. (2009). What Teachers Say and Do to Support Students' Autonomy during a Learning Activity. *J. Ed. Psych*. 98 (1), 209-218.

Sadeh, I., & Zion, M. (2012). Which Type of Inquiry Project Do High School Biology Students Prefer: Open or Guided? *Research In Science Education*, *42*(5), 831-848.

Simpson, R. D., Koballa, T. R., Oliver, J. S., & Crawley, F. E. (1994*). Research on the affective dimension of science learning*. In D. L. Gabel (Ed.), Handbook of research on science teaching and learning (pp.221–234). New York: National Science Teacher Association.

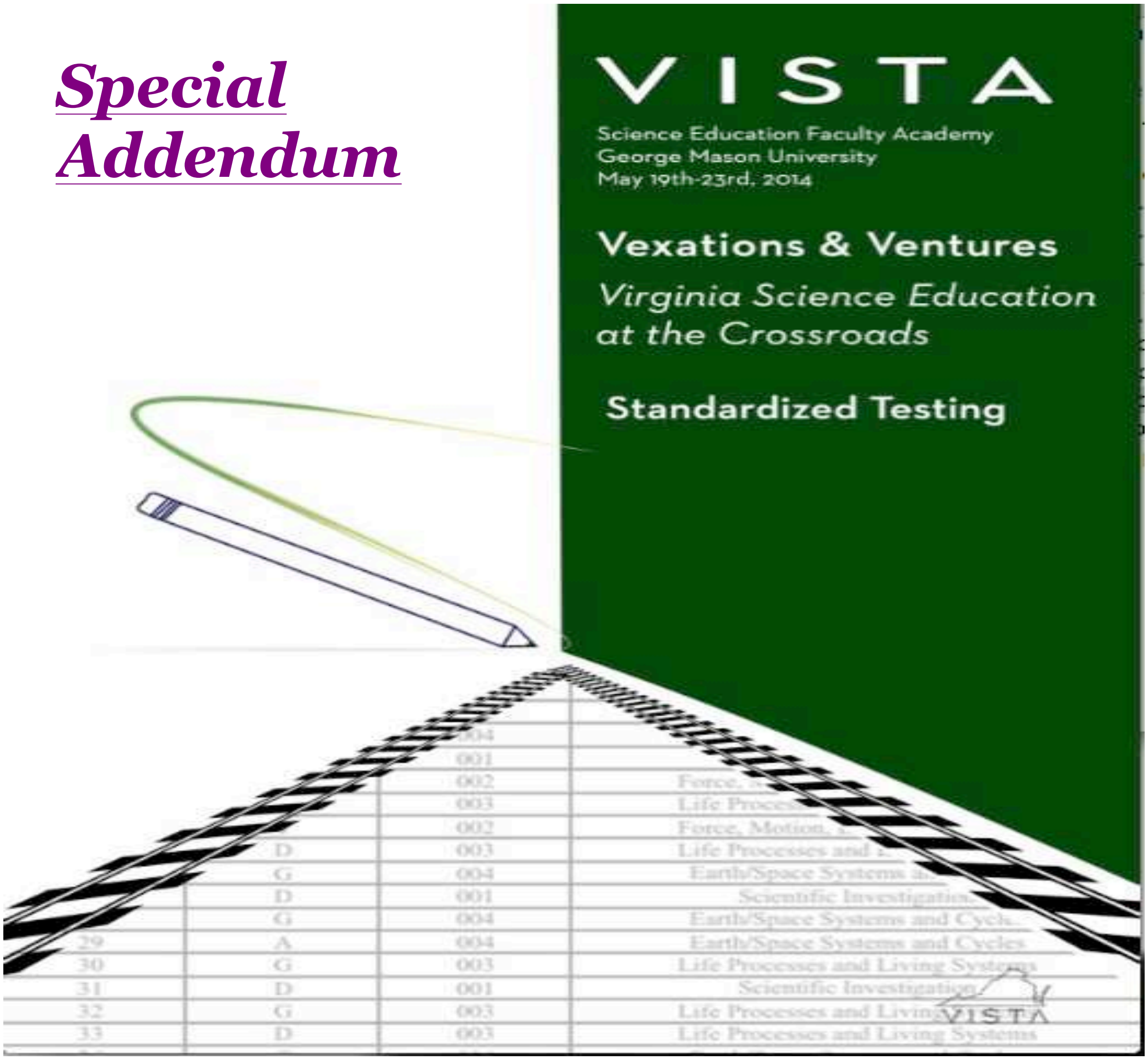
STEM Connector and My College Options Report (2012): Retrieved from http://www.stemconnector.org/sites/default/files/store/STEM-Students-STEM-Jobs-Executive-Summary.pdf

Trends in International Math and Science Study (TIMSS). (2013): Retrieved from http://www.brookings.edu/research/reports/2013/03/18-timss-pirls-scores-loveless

Udovic, D., Morris, D., Dickman, A., Postlethwait, J. & Wetherwax,P. (2002). Workshop biology: Demonstrating the effectiveness of active learning in an introductory biology course. BioScience 52 (3): 272–281.

Windschitl, M. (2003). Inquiry projects projects in science teacher education: What can investigative experiences reveal about teacher thinking and eventual classroom practice? *Science Education*, *87,* 112–143.

# ***Special Addendum***




***Vexations and Ventures, Virginia Science Education at the Crossroads: Standardized Testing***

In this section, several educators were asked to give their educated opinions and experiences with testing. The educators voice their concerns (Vexation) and their path to better understand the concern or change it (Venture). The following papers were submitted by several science education faculty across the state of Virginia who had attended the most recent Science Education Faculty Academy (SEFA) as part of the federally funded Virginia Initiative for Science Teaching and Achievement (VISTA) program. The original papers were presented using the Vexation and Ventures model of conferring developed by John Settlage, University of Connecticut and Adam Johnston, Weber State University.

## Vexation and Venture: Standardized Testing

D.L. Dickerson

### Problem

Standardized tests seldom assess students' ability to appropriately apply knowledge of scientific skills and concepts in authentic scenarios.

Arguably, one of the primary purposes of learning is to enhance one's ability to respond to phenomena in ways that are beneficial, both cognitively and physically, individually and collectively. Generally this requires application of knowledge, whether gained formally or informally. In a formal science education context, we attempt to facilitate learning of discrete concepts and skills, the connections between them, and authentic processes within the discipline. Within the classroom, science teachers regularly address such content and in order to understand the efficacy of their actions and better address student learning employ assessments.

Types of assessments used vary dramatically depending on the content, student's age and abilities, resources put towards item development, the teacher's pedagogical understandings, time for implementation, administrative support, and perceived pressure of standardized test achievement. Although this certainly is not an exhaustive list of factors impacting teachers' decisions regarding classroom assessment, they represent some of the common themes.

Within this varied landscape many informed science teachers use a variety of assessments to gain a more complete picture of student learning including vocabulary acquisition, conceptual understanding, and application in authentic scenarios. Such teachers are exercising pedagogical-content knowledge by matching the most appropriate assessment tool to the content objective and student context. The use of multiple forms of assessment regarding a single content objective is particularly powerful in finding holes in students' conceptual understandings, which can then be explicitly addressed. These teachers also gain insight into how well students can apply their understandings, often through performance-based and authentic assessments. In other words, informed science education teachers recognize the importance of knowing that their students can not only talk about the content but also do something meaningful with it. This informed approach to assessing our students' knowledge and abilities stands in contrast to traditional standardized testing.

The concerns expressed by legislators and business leaders regarding an ill-prepared workforce are often accompanied with calls for increased accountability and traditional testing. The tests implicitly shout to teachers what is valued. A discrepancy exists however between what our leaders say they value and what the tests imply they value. Often what gets tested are not the skills and understandings that would constitute a well- prepared workforce. Instead, discrete, isolated factual knowledge and memorization of vocabulary are often the focus, while conceptual understanding, critical thinking, problem solving, creativity, and argumentation fail to be adequately addressed. I do not believe our leaders are lying about their concerns for a well-prepared workforce capable of successfully competing in a global economy. So why then, would our leaders advocate for expensive policies (e.g. standardized testing) that fail to assess measures that are generally considered to be important in developing a globally competitive workforce? There are many possibilities; chief among them, failing to recognize where their expertise ends and others' begins. Tantamount to this failure is the need to maintain power and the related concern of effectively allocating limited funds. Whether a government or business leader, often a lamented but necessary critical concern is staying in the leadership role. One cannot affect change if one has no power (official or otherwise), so even well intentioned leaders often find themselves making tough choices regarding limited funds. For example, assessment items regarding conceptual understanding, critical thinking, problem solving, creativity, and argumentation are typically more difficult to produce and score. This increase in difficulty usually translates to increased costs. Leaders making decisions regarding assessment who want the best for our children but hold little understanding of pedagogical issues may view the decision to use 'cheaper' test items in order to be good stewards of the limited resources available as an easy decision without being fully aware of the consequences to the entire educational system.

## Solution (potential, partial)

Can the NGSS's *Engineering Design* provide the impetus to reform testing to include application of skills and concepts in authentic scenarios?

So what can we do to influence policy, given the classic problem that our leaders expectations too often do not match the pedagogical implications of their policy directives or funding decisions? The science education community needs to provide our leaders with opportunities and politically attractive options to do what is best for students and society. Rather than complain that they are not doing anything to help education, we need to help them help us. There are two options I will introduce in this paper: 1) cost/benefit analysis regarding testing format and enhanced local workforce and 2) leveraging the Next Generation Science Standards (NGSS) to argue for change now.

Due to considerable variability regarding economic drivers across regions, states, and even areas within states, large-scale cost-benefit analyses may not reveal data that are particularly attractive or even meaningful for local business and government leaders. Rather the science education community should attempt to make a case locally or within the state for how dollars spent in developing and implementing historically more costly authentic assessments can generate increased revenue. Science education professional organizations can help lead this effort by organizing the collection and dissemination of such information. The science education community, as a whole, will need to better attend to and engage in policy discussions. Science education professional organizations are best positioned to take the lead in helping the community become more proactive rather than reactive. Additionally, the role of authentic assessment as a valid tool for measuring meaningful workforce skills and understandings may lead to cost saving opportunities to engage students in public service efforts. For example, if students were able to demonstrate their knowledge through a public works effort there may exist not only a potential to cut testing costs, but also an opportunity to save tax payer money regarding labor associated with that public work.

With the advent and implementation of the NGSS, curriculum revision cycles will begin to address standards explicitly related to design process and engineering. State standards will emerge and discussions of assessment will ensue. The nature of the Practices, Disciplinary Core Ideas, and Crosscutting Concepts within Engineering Design focus heavily on complex processes and application of science and mathematics concepts. The NGSS may serve as a catalyst for not only curricular change, but also change in assessment. This change will not occur however without explicit, concise arguments that focus on elements that resonate with our leaders. The science education community has an opportunity that it has not had since the mid 1990's to make significant changes to the ways in which we conduct assessment and the NGSS has the language to help facilitate this move. The Performance Expectations sections have both the initial appearance and the philosophical underpinnings to support authentic and performance-based state- administered assessments. Certainly there are other, most likely better, solutions that exist. These two discussed here represent broad ideas that can be used as a springboard to further idea development and, it is my sincere hope, action.

**Does Time On Task Determine What We Learn?**

Tricia Easterling

If crossroads are indeed a juncture where paths cross and minds can meet briefly, they also serve as signposts. Crossroads are waypoints where travelers stop, check their bearings, and consider in what direction they would like to continue. House Bill 930 provides an opportunity for local school districts to determine how *not* testing science content will affect the teaching and learning of science in their communities.

HB states "The Standards of Learning assessments administered to students in Grades 3 through 8 shall not exceed (a) reading and mathematics in Grades 3 and 4…" (Virginia, LIS, 2014).

**If Not In Science, Then When?**

It is my belief that if we do not test as much science, educators will not teach as much science. If we do not teach as much science, we will not learn as much science. If we do not learn what science is and understand how it shapes our thinking, students will not reason as effectively. If students do not learn science content or understand how we use an objective approach to make sense of the world, it is my belief students will increasingly exhibit self-centered habits of mind. For example, without science developing curiosity, will students continually seek to know and understand? Or will students become intellectually arrogant, thinking they already know it all?

Science class offers an opportunity for students to recognize and respect a logical and systematic approach to solving problems and learn new information. Where will students learn to be methodical and organized? What venue will teach students to observe? Where will students gain insight into the relationship between noticing and learning? When will students learn to measure or appreciate accuracy?

In science classes we teach students to interpret all available data. Without a science class will students become increasingly indifferent to new information? Instead of basing their judgment on evidence, will students be more emotional in their decision-making? What class will teach students to be comfortable with ambiguity? What class will teach them to be resilient until results confirm or deny their thinking? Regarding natural phenomena, students are encouraged to observe, to make inferences, and notice patterns. Where will these strategies be featured if not in the science classroom? If hands-on learning happens in science, what will happen if that concrete experience is taken away? Will we grow increasingly dependent on teaching using only images and text?

Not only do we need to teach students, we need to help them learn. While these two ideas may appear redundant, they differ considerably. Teaching has a valuable and necessary role in directly sharing ideas and strategies and providing a logical sequence of ideas. But to help a student learn implies an intimacy; Getting insight into their understanding can only come from spending time with them. The deeper habit of thinking and reasoning through a problem, gaining resilience and grit take place with experience, with repeated exposure, with practice. Depth can only be gained by going deep. Experience changes the brain; “[T]he repeated experiences kids have alter the actual architecture of the brain” (Bryson, 2014). If we do not dedicate time to teach and learn science in the earliest of grades, how will we ignite students’ curiosity? We won’t. The joy of experimenting will be slowly extinguished in earlier grades.

As Einstein suggested, science is a venue that can teach students how to think (1943). Science methods professors can make students aware of their habits of mind, if you will; demonstrating strategies that support students learning on their own. If science is not going to be taught to students, they need to learn how to teach themselves. In science class teachers can model and practice the habits of mind to reach the ultimate goal of an education: to become intellectually independent.

# VISTA Science Education Faculty Academy (SEFA):

## Vexations and Ventures of Standardized Testing

Harold A. Geller

Not long ago, I received an e-mail from a colleague, now retired, wherein he noted that he had discovered an interesting webpage titled "30 Brilliant Test Answers from Smart-ass Kids." As it happens, I recognized some of those amusing responses from kids taking exams, but some were new to me. Then it occurred to me that I could use this as a starting point in my V&V response for the SEFA assignment. How is it connected? Well, most of these humorous responses were given by students on a type of standardized examination, which I will define shortly. However, instead of giving a typical response, these students had some very clever answers. Now, as the title of the online article alluded to, some of these responses were from students who didn't know the answer and were just being wise-asses. But some showed a creativity that standardized tests just simply cannot accommodate. Some of these kids are far brighter than any standardized test would ever note. This is one of the problems with standardized testing.

**Vexation**

One of the problems I have with standardized tests is that they are looking for a specific answer to each question. Those students who think outside the box are ultimately penalized and may even be labeled as being unintelligent; however they may be quite brilliant. A frustration I have with any discussion of standardized testing is that there always appears to be a new issue. So I wondered how far back the era of standardized testing actually went. In one reference I found that standardized testing was used to evaluate teachers in the United Kingdom in 1710 (Harris et al., 2011). "Wow," I thought, "so arguments about standardized tests go back at least

300 years!" But then in another reference I discovered that standardized testing goes back a few thousand years. In fact, standardized tests were used by Chinese emperors to determine who would be best qualified for civil servant positions in the empire (Garrison, 2009). So, this idea of standardized testing is not exactly a new concept. Why are we still arguing over standardized testing after more than a few thousand years?

In 1992, the Office of Technology Assessment wrote a report titled *Testing in American Schools: Asking the Right Questions.* The report concluded that the development of standardized tests in the $19^{th}$ century "was born in the minds of individuals already convinced that education was substandard in quality." In fact, the report concluded that "tests are often administered not just to discover how well schools or kids are doing, but rather to obtain external confirmation——validation——of the hypothesis that they are not doing well at all." (Office of Technology Assessment, 1992)

Garrison is nowhere near the only academic who believes that standardized tests are not exactly what they are made out to be. In *The Myths of Standardized Tests: Why They Don't Tell You What You Think They Do*, Phillip Harris, Bruce M. Smith, and Joan Harris stated early in the introduction "our schools are under attack" (p.1). The authors point to the latest barrage of tests instituted under No Child Left Behind and continuing into the Obama administration as being indicative of a "misguided reliance on test scores as the primary measure of success for students, teachers, and the education system as a whole" (Harris, Smith, & Harris, p. 1). At one point, these same authors conclude that "no matter how you slice it, when high stakes are attached to tests, lots of test preparation will take place" (Harris, Smith & Harris, p. 95). I would say that this would put a damper on the true enjoyment of learning and teaching science.

**Venture**

In my vexation above I have given but a sample of evidence selected from a wealth of additional documentation against the many myths regarding standardized tests. But what is the solution?

Of course, there is no single solution to the issues raised by standardized testing. However if I were master of this universe I would start by eliminating standardized testing for all the lower grades. Imagine you went to an elementary school where there was no standardized testing. What would it be like? Well, at least you would not hear about teachers "teaching to the test." That is a start.

The only testing I see as necessary begins in high school, when students need to demonstrate to colleges that they have actually learned something in Grades K-12. Standardized tests can begin to be disseminated in ninth grade. In fact, these tests should actually be the very same tests that the students take in their senior year in high school. Yes, the students need to be carefully taught that they will be entering the twilight zone of standardized tests, and that they should just get used to taking such examinations. In ninth grade, at the start of the term, students should see examinations in American history, world history, English composition, English literature, geology, biology, chemistry, and yes, physics. These same examinations should then be retaken by the students each June after completing Grades 10, 11, and 12. That's all there is to it. No breaking up the curriculum at hand by preparation for the examination at any time, just give all the exams the last week of classes. Such tests would be utilized by colleges to augment the standardized tests of all standardized tests, the ACTs and SATs. Are we ever going to get rid of those despite the lack of evidence that they are great predictors of academic success in college (Kohn, 2000)? No, these college gatekeepers are corporations of truly epic size and power, not to mention wealth. They will always be in existence, for better or for worse, regardless of their true value.

Before concluding, I would like to return to that web page with the so-called smart-ass responses. All but one of the examples provided by the boredpanda webpage are open-ended questions or fill-in the blank. The one multiple-choice question highlighted shows that the student was so upset by the question and its choices that each student wrote in his or her own choice. This multiple-choice question has three associated graphics. In the first diagram you see three giraffes, two taller than the third. In the second diagram you see the two tallest giraffes still standing, but the third, the shorter giraffe, you see lying down on the ground. In the third diagram, you see the two tallest giraffes standing and eating leaves while the third giraffe is only a pile of bones. Now the "correct" choice of the four provided choices for this question is Choice A. After all, the question starts with "The diagram below illustrates" and the choices all complete the sentence. Choice A is Lamarck's theory of evolution. However, in this case, the student added Choice E. Associated with Choice E is the statement "Giraffes are heartless creatures." In actuality, given the illustrations provided, not realizing that it was meant to be a long-time series of illustrations, the student's response is certainly a valid one. Of course, adding one's own choice alternative violates test taking, and the student misinterprets the sequence of illustrations because of some kind of lack of understanding of the intended concept being questioned. Nonetheless, the wrong answer should not be interpreted to mean that the student doesn't know anything about evolution.

One of my favorite open-ended questions on the webpage is "Miranda can't see anything when she looks down her microscope. Suggest one reason why not." The student response was "She is blind" (30 Brilliant Test Answers). Well, if Miranda were blind, the statement would indeed be true, so why mark the student as being wrong? It is because the teacher was expecting a particular response, but the student was thinking differently, not incorrectly.

What about the question, "Where was the American Declaration of Independence signed?" The student wrote, "At the bottom" (30 Brilliant Test Answers). Now, who is to blame for this wrong answer? It is just as reasonable to state that the question itself is not worded well enough. It is too vague, so "at the bottom" is not really a wrong answer. However the student gets no credit. Does this mean the student is not smart? I do not think so. The student has simply not learned yet how to conform to the test maker's way of thinking.

There is an abundance of historical evidence, going back thousands of years, to support the case that we will always have standardized tests. The companies that prepare them are too big to fail, and we simply are more interested in our own political interests rather than what is best for the individual student or society at large.

**References**

*30 Brilliant Test Answers From Smartass Kids* [Web log post]. (2014, 4 April). Retrieved from http://www.boredpanda.com/funny-test-answers-smartass-kids/

Garrison, M.J. (2009). *A Measure of Failure: The Political Origins of Standardized Testing.* Albany, NY: State University of New York Press.

Harris, P., B.M. Smith and J. Harris (2011). *The Myths of Standardized Tests: Why They Don't Tell You What You Think They Do*. New York: Rowman & Littlefield Publishers, Inc.

Kohn, A. (2000). *The Case Against Standardized Testing: Raising the Scores, Ruining the Schools.* Portsmouth, NH: Heinemann.

Office of Technology Assessment (1992). *Testing in American schools: Asking the right questions*. Washington, DC: U.S. Congress.

Ravitch, D. (2010). *The Death and Life of the Great American School System: How Testing and Choice are Undermining Education*. New York: Basic Books.

# Standardized Testing

Marilyn Lanier, Radford University

## Vexation

There is likely no other educational topic that generates as much discussion as standardized testing. The Virginia Standards of Learning (SOL) were designed to test and measure student progress while ensuring accountability for student achievement. The underlying question continues to be what the SOLs tell us about how students learn science and how this understanding can help us as science educators. It can become frustrating knowing that as a teacher, you could be judged on the performance of your students on this one criterion test and not on your teaching pedagogy and skills. As a result of this increasingly negative stigma on the teaching profession, other areas are also being impacted, including a fall in first year teacher retention rates and decreasing interest from prospective future teacher candidates in entering the career field. It may even deter experienced in-service teachers from accepting assignments in high-risk or high-needs school districts (Fair Testing, 2012).

School systems have begun to face the very real consequences of state-mandated standards, testing, and accountability. The use of such has made *content standards* a dirty word in the minds of teachers and students alike. Reaction to this has been as varied as it has been intense. Policymakers face continued frustration over obstacles to implementation and limited progress toward closing achievement gaps, while press releases from the White House and United States Department of Education cite narrowing the gaps and improved performance on National Assessment of Educational Progress indicators (Fair Testing, 2012). On the other hand, teachers lament the loss of autonomy and report having to narrow curriculum to the tested standards, particularly in relation to schools serving large populations of minority students. The standards set minimum requirements in each grade level (kindergarten through twelfth grade) that specify what teachers need to teach and students need to learn in order to meet the nation's call to action to achieve scientific literacy for all learners (National Research Council, 1996). In 1995 the Virginia Board of Education encouraged school systems to adopt these prescribed standards, the Virginia SOLs, to assure enrichment of the local curricula that will meet the needs of all of the students who are serviced in the area. But the needs of all students are not being met by this attempt at science educational reform. These Virginia standards-based assessments have been inappropriately used as the main gradient tool, or assessment mechanism, to evaluate what students know in various subjects at various grade levels since the enactment of No Child Left Behind (NCLB) in 2002.

The standards appear to be increasingly controlling the design and development of curriculum within our schools each year. Curriculum, assessments, and professional development programs are aligned to the standards. The standards-based reform movement calls for clear, measurable standards for all students. Expectations are raised for all students' performance. Rather than norm-referenced rankings, the overall performance of students are expected to be raised. Because the standards present specific, definable goals, I believe this is a realistic expectation. With qualified classroom teachers leading the instruction and with adequate supplies, students should rise to the expectation. To better match the science SOLs of 1995 and 2010, many schools have realigned their curriculum. Teachers have also refocused their instructional programs to accommodate student learning abilities. The assessment component of the reforms, school accreditation, and school report cards keep educators and parents alerted to what is expected in the classroom.

However, pressure has increased upon teachers and school administrators to ensure that students are consistently improving. As a result, much of the school curriculum is geared to match the items on the SOL test, and in an effort to meet state standardization requirements, teachers find themselves restricted to repetitive teaching, i.e., “teaching to the test,” which primarily involves drilling students to recall rehearsed information rather than engaging in reflective instruction that involves science explorations and investigations (Haberman, 1991). In many cases the depth and rigor of a content area suffers, resulting in more test prep than content instruction. In these cases, students from low-income and minority-group backgrounds, English language learners (ELL), and students with disabilities are more likely to be inaccurately evaluated.

A successful teacher must be able to display mastery of the content, manage his or her classroom, while still encouraging spontaneous interactions that promote learning, earning the trust and respect of her students. Teachers must also be willing to engage in the reciprocal sharing of knowledge and experience to reinforce the established curriculum. If education is to be viewed as a pathway out of the pedagogy of poverty and as promoting student achievement, then there is a significant need to implement pedagogical methodologies that are culturally responsive and connect to student lives (Lanier & Glasson, 2014). Simply teaching to the test, as asserted by Haberman, may further alienate urban students from the actual practice of doing science within the classroom and making the necessary connections for minority and urban students alike.

Effective evaluation methods provide useful information to teachers and students. Because scores are based mostly on multiple-choice and short-answer questions, research supports the notion that most teachers do not find scores from standardized tests very useful

(Fair Testing, 2012). Nor do teachers agree that they measure what students have learned, but rather, only what they can recall at the time of testing.

In conclusion, the Virginia standardized achievement tests are useful in that they do supply evidence used to compare our students with other students nationally, using a norm-referenced analysis of students' knowledge and skills. But a classroom teacher performs teaching and standardized achievement tests scores are but one product of student performance and should not be used to evaluate the quality of teaching or education. Standardized test scores are simply insufficient in that regard.

**Venture**

While there are many arguments against standardized testing, we must consider alternative approaches to measuring student achievement and abilities. Standardized tests can be one part of a comprehensive assessment system but should not be the sole or major assessment tool utilized to measure student achievement or teacher effectiveness.

Standardized curriculum is based upon a set of prescribed statements that set reasonable targets and expectations with respect to content, processes, and skills as determined by the assigned administrative team of teachers, educators, and stake-holders. The early national curriculum standardization reform efforts of the 1960s were unsuccessful because they failed to recognize that in order to promote change in the classroom, one must begin with the classroom teacher—that their beliefs and pedagogical practices would inevitably impact student learning. Reforms were also ineffective because of poor teacher preparation, as well as the fact that instructional strategies were not making adequate connections to the knowledge students brought from their home cultures. As a result, American students score lower internationally in mathematics and science as compared to their Japanese and Chinese counterparts (OECD, 2010).

Consideration is given to the pedagogical strategies used in other countries. For example, Finland, one of the top scoring international countries, utilizes practices such as good teacher observation, documentation of student work, and performance-based assessment, all of which are well thought out and are applied when directly evaluating real learning tasks (Fair Testing, 2012). This type of information provides useful materials for teachers, parents, and the public to reflect upon when measuring student achievement. Remediation efforts—including prevention and intervention activities—have never been more important. Funds are also needed to support the training of teachers in remedial techniques or strategies.

In summation, I concur with the current research. More authentic assessment should be incorporated in the summative evaluation of students. Other examples would include a collection of student work in portfolios, e-portfolios, and written explanations such as essays and projects. This would appear to be more useful and accurate for measuring achievement (as well as providing more information for teaching) than multiple-choice achievement tests. In addition, I further posit students should have a pre- and post SOL test for each grade level currently assigned to be tested. With the pre- and posttest, students and teachers can assess and address the contextual needs and plan appropriate strategies for success. This could ease the tension and stress of testing and diminish the labeling and shame of not passing the test. Finally, consideration would need to be given to safeguards to ensure that, with multiple-choice exams, student characteristics such as race, class, gender, linguistic, or other cultural biases do not affect evaluation. Ironically, because teachers of the higher scoring nations do not focus on teaching to multiple-choice and short-answer tests, their students score higher on international exams. Achievement and learning should be functional and progressive to self and society (Dewey, 1964), and the teacher should be a skilled facilitator in fostering such an educational paradigm. Therefore, these strategies may be worth exploring further.

**References**

Dewey, J. (1964). *John Dewey on education: Selected works* (R. D. Archambault, Ed.). New York: Random House.

Fair Testing (2012). How standardized testing damages education. Retrieved 10 April 2014 from http://fairtest.org/how-standardized-testing-damages-education-pdf

Haberman, M. (1991). *The pedagogy of poverty versus good teaching*. Phi Delta Kappan, 73, 290-294.

Lanier, M., & Glasson, G.E. (2014). Investigating strategies for enhancing strategies of urban African American students in middle school science classrooms. *Insights on Learning Disabilities, 11*(1), 9-32.

National Research Council (1996). *National science education standards.* Washington DC: National Academies Press.

No Child left Behind (2001). The White House. Retrieved November 13, 2006 from http://www.whitehouse.gov/news/reports/no-child-left-behind.html

OECD (2010). Shanghai and Hong Kong: Two district examples of education reform in China. Retrieved 14 March 2012 from http://www.oecd.org/dataoecd/34/45/46581016.pdf

**All Politics are Local: How Standardized Testing is Playing out in Virginia**

Juanita Jo Matkins

In January 2012, when a Virginia state senator introduced a bill to eliminate the third grade Standards of Learning (SOL) test for science at the onset of the 2012 session of the General Assembly (lis.virginia.gov), I was President of the Virginia Association of Science Teachers (VAST). We learned about this bill at our first board meeting of the year, and the board decided to take quick action to empower a group of board members to advocate in opposition to this bill. This resolve was based on the prediction held with high confidence by board members that the unintended outcome of this bill's passing would be a dramatic decrease in time spent on elementary science instruction. The VAST Advocacy Committee members, the Virginia Science Education Leadership Association (VSELA), and the Virginia Mathematics and Science Coalition (VMSC) rallied their various membership constituencies and launched a successful campaign in opposition to the bill (SB185/Miller). The committees were successful because the bill was tabled and not acted upon in the 2012 session.

Critical in the reasoning behind VAST's opposition to SB185 were several published studies about (a) curriculum narrowing (Berliner, 2009; King & Zucker, 2006) resulting from standardized testing, (b) narrowing specific to science instruction in elementary schools (Srikantaiah, 2009), and (c) impact on the science, technology, engineering, and mathematics (STEM) pipeline and workforce of early experiences in science (Jones, Taylor, & Forrester, 2011). These results were summarized, cited, and repeated in communications with legislators and with member constituencies in the science education organizations. Again and again, we emphasized that VAST and the other organizations were not in favor of more tests for elementary children; Instead, our efforts were focused on supporting science instruction in elementary schools in Virginia.

In January 2013, a set of three bills were brought to the Virginia House and Senate (SB1364/Miller, SJ306/Miller, and HB2144/Landes) along with an action plan set forward by then Governor McDonnell to support on-grade level reading achievement by third grade by eliminating third grade end-of-year assessments in science and social studies, with the stated intent that time gained from teachers not having to prepare for the science and social studies SOL assessments would then be diverted to time spent in reading instruction with students. VAST and the other organizations again lobbied vigorously in the legislature, and ultimately the judgment was made that a version of at least one of these bills was going to pass—probably the Governor's bill (HB2144/Landes)—and negotiations were initiated to put science educators in the conversation about how this bill would be enacted. The bill passed that year, stipulating that only schools who were not achieving a 75% pass rate in reading in third grade would qualify for the SOL assessment waiver, provided they also met several other conditions. The schools must have a full-time reading specialist in the school, commit to 30 minutes daily instruction in science and social studies, and administer a summative test in every waived subject. This bill went into effect July 2013, and waivers are in effect for the first time in May 2014. Thirty-eight schools met the pass rate qualification at the time of the bill's passing, and results from this legislation are unclear at this time.

In January 2014, it seemed the floodgates opened in terms of legislation to reduce SOL end-of-course tests in Virginia. Seven bills were introduced affecting the number of standardized tests administered in Grades K-12: HB447/Minchew, HB498/Krupicka, HB640/Campbell, HB930/Greason, SB270/Miller, SB306//Deeds, and SB325/Miller. Senator Miller of Hampton, the patron of the 2012 and 2013 bills, was joined by several other legislators across the political and legislative aisles. Advocacy efforts of VAST, the VMSC, VSELA, and several other organizations were launched to stem possible damage from several well-meaning but potentially damaging aspects of these bills, including the elimination of several high school science SOL tests without consideration of how that move would affect verification of credits—a step required for graduation from high school. What emerged from this flurry of bills was one bill that consolidated several of the other bills, HB930, and that bill was where advocacy efforts were concentrated. In the midst of the legislative session, newly elected Governor McAuliffe's Secretary of Education, Anne Holton, called a meeting with Delegate Greason and Superintendent of Education Pat Wright. Following this meeting revisions to the bill were introduced. The key change that emerged from that meeting was the addition of a SOL Innovation Committee. This committee was tasked with oversight of any efforts to eliminate tests and also oversight for the development and implementation of alternate assessments. The science SOL test for third grade was eliminated. The fifth and eighth grade science SOL tests as well as all high school science tests were retained. Key provisions from the 2013 legislation were repeated in HB930, including requirements that content not tested must still be taught, assessed, and reported. Key areas left undecided by this bill that concerned science educators in Virginia included (a) specific participation of science educators in the SOL Innovation Committee, (b) the nature and use of alternate assessments to be used to show science achievement when tests were waived or eliminated, (c) oversight for requirements of instructional time spent on teaching science, and (d) consequences when requirements were not met. That is where things stand at this point.

During the legislative session, I was also teaching a class for VISTA. The students were middle and secondary science teachers in Tidewater school divisions. A conversation in class revealed that at least two of the teachers had devoted a large amount of instructional time—28 class periods out of 180 total—to standardized testing or test preparation, relating to either the district-mandated benchmark tests or the state-mandated SOL test. One-third of their instructional time was given over to testing at the direction of their administrators. This is one of the reasons I was pleased by the passing of HB930.

**Vexation.**

My vexation is finding satisfying answers to the question "How should we measure the outcomes we, as science educators and as members of families and citizens, want from K-12 science education?" Currently the most popular way is through testing all students with tests that are easily scored and whose results are easily compared. Conley and Darling-Hammond (2013) advocated for a continuum of assessments tied to curriculum and to the kinds of learning and practice needed for college and career readiness. What can those look like in Virginia? And how do we move away from the negative impacts on instruction we've seen without exchanging these impacts for equivalent or even more damaging choices? Recently our newly appointed Superintendent of Education, Dr. Steve Staples, offered the idea of using science fair projects as an alternative assessment for science achievement. That prospect concerns me greatly, since I have directed science fairs for a Central Virginia school district for several years and have found schools unable to compensate for discrepancies of parental support. My fear is that such an option may only increase the achievement gap related to economic status.

**Venture.**

Science educators must continue to make their voices heard with legislators and administrators in Virginia. This should include not only scientists and university science educators but also teachers of science. The SOL Innovation Committee should include a science educator who has knowledge of alternative assessment and probably who has experience with science fairs. Dr. Staples will most likely be a powerful voice in future education policy discussions in Virginia, at least for the near term. And we need to decide what we think are valid and viable assessments for science achievement. Do we go with the culminating projects from problem-based learning curricula, with science fair and Virginia Junior Academy of Science (VJAS) research projects, or with a menu of options? Most importantly though, we need to be in on the conversation.

**References**

Berliner, D.C. (2009). MCLB (Much curriculum left behind): A U.S. calamity in the making, *The Educational Forum, 73:*4, 284-296.

Conley, D.T., & Darling-Hammond, L. (2013). *Creating systems of assessment for deeper learning.* Stanford, CA: Stanford Center for Opportunity Policy in Education.

Legislative information on all bills referenced. Retrieved from *http://lis.virginia.gov/*. April 14, 2014.

Jones, G., Taylor, A., & Forrester, J. H. (2011). Developing a scientist: A retrospective look, *International Journal of Science Education, 33:12*, 1653-1673.

King, K.V., &Zucker, S. (2008). *Policy report: Curriculum narrowing.* Washington, DC: Pearson Education, Inc.

Srikantaiah, D. (2009). *How state and federal accountability policies have influenced curriculum in three states.* Washington, DC: Center on Education Policy.

**Vexation and Venture: Might There Be a Way Out of the Testing Loop?**

Jacqueline T. McDonnough

I can safely say that American public education is in the throes of a testing frenzy. We test children to rate schools' adherence to national guidelines. We test children to evaluate teacher effectiveness. We test children to make sure that they will pass their end of course tests. Some of the tests are created by national textbook companies, others by teams of district staff from test banks, and yet others are created by teacher teams. All of this testing is being done at a cost. The cost can be measured in lost instructional time, teacher and student anxiety, and the diversion of instructional resources toward test preparation and administration. There are hidden costs that oftentimes get lost—these include the erosion of teacher autonomy and demoralization (Ravitch, 2014) and students' dissatisfaction with their education experience. So what are the roots of this current testing crisis?

The roots of this crisis sprouted in the Commonwealth of Virginia at a 1989 meeting of an education summit convened by then president George Bush. At that event Albert Shanka, President of the American Federation of Teachers, urged the president to create a national system of standards and assessment. The call for a nationwide system of standards and assessment grew out of a desire to increase student achievement (McDonnough, 2002). The outcome of this meeting was policies that fueled the creation of state-level accountability initiatives. Embedded in many of those initiatives were policies that set benchmarks for increasing student proficiency in mathematics and reading with penalties for schools that fell short. The crowning achievement of the accountability movement was the 2001 No Child Left Behind (NCLB) legislation. These policies were heralded as a game changer in our quest to close the racial achievement gap and a testament of cooperation between both sides of the House in Congress. In Virginia we used the

Standards of Learning Assessment to meet requirements set forth in this legislation. Between 2002 and 2009 we saw major shifts in schools towards focusing on reading, writing, and mathematics instruction and a lessening of focus on physical education, history, and science instruction. In my opinion this trend was more pronounced in schools with large populations of racial minority and English Language Learner (ELL) students. In 2010 the Obama administration signed into law policies that were intended to correct some of the negative effects of NCLB, but by all measures this seems to have worsened the problem. These Race to Top policies included provisions to focus more on student improvement and less on accountability punishments, incentives to keep students in school, appropriate assessment of ELL and racial minority students, and a new evaluation program for teachers and educators (No Child Left Behind, n.d.).

**Vexation**

We have painted ourselves in a corner in terms of high-stakes student assessment. Reliance on one form of assessment to measure student learning and, in turn, teacher efficacy has resulted in a runaway testing culture that threatens to totally undermine the purpose of education. How can we mitigate the impact standardized testing has had without sacrificing the good that comes with setting standards and assessing student learning?

**Venture**

How can teachers stay true to their mission of preparing children for a complex, ever-changing world while adhering to the standards and taking into consideration the ever-increasing diversity of students in their classrooms? I propose a two-pronged approach. Let's hit the reset button on high-stakes testing and assess the impact that it has had on students, that is, curriculum narrowing and decreased motivation to name a few. We should also look at how this singular focus on testing has eroded teachers' capacity to use research-based pedagogy in their instruction and teacher morale. As education researchers we have the opportunity to study these issues within the context of our jobs. A well-structured research study probing these issues may bring some clarity to the issue and point a way forward as we struggle to find a way out of the testing darkness. In addition to research on the impact of high-stakes testing on teachers, we can squarely confront curriculum narrowing and decreased student motivation by looking at pedagogical fixes.

The argument against reliance on high-stakes testing in no way hinders adherence to standards. We need standards to ensure all children, no matter their zip code, have access to a high-quality education. But standards, though necessary, are not sufficient. Our students represent a diverse mix of cultures, and the curriculum and pedagogies we use in our teaching should reflect that. As science educators we have a responsibility to equip our preservice teachers and the in-service teachers we train with the tools to deliver science content that is both standards based and culturally relevant. I propose using a curriculum framework designed by Education Northwest called *Culturally Responsive Standards Based Teaching* (2011). Their framework uses the culture and funds of knowledge students bring to school by integrating it with curriculum standards to produce engaging, student-centered instruction. Using this approach values both the students' culture (however that is defined) and standards set by education stakeholders.

Determining what aspects of students' culture to focus on will take some work on the part of the teacher. Teachers have to first recognize that culture is an essential part of all individuals lived experience including her or his own. Then they have to acknowledge the importance of including aspects their students' cultural landscape in their teaching. How can they do this, you may ask? They have to know their students and their communities. One step could be conducting surveys at the beginning of the year that try to get at unique celebrations or families histories and incorporating that information into interactions in the classroom. Another may be researching different cultural ways of thinking about science and incorporating those ideas into science instruction. These are but a few ways to start collecting information that would enable teachers to integrate cultural aspects into their instruction.

**References**

McDonnough, J. T. (2002). Implications of reported use of constructivism with diverse learners. Unpublished doctoral dissertation. The University of Virginia.

No Child Left Behind. (n.d.). In *Wikipedia.* Retried May 5, 2014 from http://en.wikipedia.org/wiki/No_Child_Left_Behind_Act#Proposals_for_reform

Ravitch, D. (2014). High-stakes testing. Retrieved May 5th from http://www.washingtonpost.com/wp-srv/special/opinions/outlook/spring-cleaning/high-stakes-tests.html

Saifer, S., Edwards, K., Ellis, D., Ko, L. & Stuczynski, A. (2011). *Culturally responsive standards based teaching: Classroom to the community and back 2nd ed.* Thousand Oaks, CA: Corwin.

# Less Emphasis on SOL Outcomes, More Emphasis on Curriculum

Melissa Rhoten

## Vexation

I have been teaching chemistry at the college level for 14 years. Thus, I have witnessed first-hand the transition from voluntary standardized testing to state-mandated Standards of Learning (SOL) testing. In my opinion the students I see in my classes now are less prepared, less motivated, and less inquisitive than the students I had at the start of my career. I attribute this in part to SOL testing. Too much emphasis is placed on how students score on these tests, meaning that teachers must focus solely on content while leaving out important process components (information processing, critical thinking, problem solving, teamwork, communication, management, and assessment). The majority of my current students are not curious about science, which I think stems directly from the "teach to the test" mentality that many teachers must adopt to enable their students to score well on SOL tests. Thomas Armstrong stated that "the adventure of learning, the wonder of nature and culture, the richness of the human experience, and the delight of acquiring new abilities all seem to have been abandoned or severely curtailed in the classroom…" (Armstrong 2006). Armstrong (2006) goes on to say that our current educational crisis is the direct result of the No Child Left Behind (NCLB) Act. This legislation mandates annual testing of students in reading, mathematics, and science and requires that schools make "adequate yearly progress (AYP) incrementally on a year-by-year basis until *all* students reach 100% proficiency in these areas by the year 2014. Failure of a school to maintain AYP will result in penalties for the school…and the eventual placement of a school on probation leading to possible government or commercial takeover" (Armstrong 2006).

Many educators have realized that we are producing scores of students with a "tell me what I need to know for the test" attitude. This has led to the development of new and innovative pedagogies that attempt to intermingle both content knowledge and process skills. Some of these innovations include Problem Based Learning (PBL), Process-Oriented Guided Inquiry Learning (POGIL), Flipping the Classroom, and Meaningful Watershed Educational Experience (MWEE). In a 2004 study, Berube developed a "Comprehensive" eighth grade science SOL test and administered it to students in two different learning environments. Approximately half of the students tested were taught by teachers who identified themselves as constructivist teachers, whereas the other students' teachers identified themselves as mixed teachers. Those teachers self-identifying as mixed used both traditional and constructivist pedagogies. The Comprehensive SOL test used in this study was the same SOL test the students had taken a week earlier, with one stark difference. After each multiple-choice question, students were asked to justify their answers using short-answer or essay format. Berube hypothesized that students in constructivist classrooms would score higher on the normal and the comprehensive SOL tests compared to students taught by mixed teachers. Statistical analysis showed exactly the opposite:

> Scores were not higher for constructivist teachers; the more mixed and traditional…teachers produced the students with the higher SOL scores. But many of those students still failed my comprehension test in high numbers. Surprisingly, the constructivist teachers, whom I thought would produce the highest scores all around, produced high pass rates on the SOL but the lowest pass rates on the comprehension test. Students could answer the simpler multiple-choice questions but could neither defend nor explain their answers, which require higher-level skills than memorization. The constructivist teaching style (high-level teaching, according to Bloom's taxonomy) that some teachers employed during most of the year was essentially useless in preparing students for the high-stakes test. (Berube, 2004)

The findings here led the author to postulate the following questions: "… do high-stakes tests cancel out any form of higher-level teaching and learning? Do teachers who previously taught at high levels resort to 'teaching to the test' during assessment crunch time at the end of the year?" (Berube 2004).

While I fully support innovation in the classroom, I do feel that less emphasis on SOL outcomes would allow teachers the freedom to teach students content and process skill that are vitally needed. During my own elementary, secondary, and post-secondary education, none of my teachers or professors used any of these techniques, yet my peers and I learned the content and how to problem solve, think critically and creatively, and collaborate. I was able to dig deep into a topic, wrestle with it, and glean additional information about the topic not presented to me during class. The students I see now either cannot do this or do not want to do this.

**Venture**

Currently I am serving on Longwood's Core Curriculum Committee to revamp our general education program. Time and time again I have heard colleagues quote studies saying that employers are not "that concerned with a student's major" but rather they are looking for students that can "think critically, communicate effectively, and solve problems creatively." If this is what we truly value in education, then there has to be a shift in the way we measure students' abilities in these areas. Of course, students must learn content and we must ensure that they are well-versed in their subject areas, but does this mean that we need to place so much emphasis on standardized testing? According to Diane Ravitch, "Our schools will not improve if we value only what tests measure…What is tested may ultimately be less important that what is untested..." (Ravitch ,2010). Perhaps we need to step back and evaluate what we really think a quality education entails, and then develop a means of assessing those goals. Berube stated:

> The problem is not the standards but, rather, how the standards are assessed. I would have been unable to teach without the Virginia Standards of Learning. They gave me a wonderful roadmap that I coordinated with my curriculum guide and the district's guidelines. But something went terribly wrong when my former state (Virginia) began measuring the standards with multiple-choice, high-stakes tests. These tests hold teachers' and administrators' creativity hostage and threaten job security and professional contentment. In addition, the tests hardly prove that students have learned anything other than rote memorization of facts. (Berube 2006).

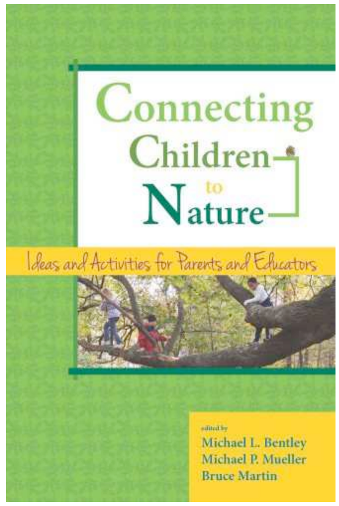
Longwood's Core Curriculum Committee is just in the early stages of reforming our general education curriculum. Based on preliminary conversations with various stakeholders, I can only assume that the new program will be more integrated and interdisciplinary, perhaps having a curriculum where a fair number of the required courses are woven together to address a particular theme like "Addressing World Health Issues" or "Art in Society." This would help students understand the need for multiple perspectives and the relationships and interconnectedness of all subjects when facing real world issues and challenges. New courses will need to be created to fit into this framework with the goal of producing students not only with content knowledge, but with the ability to communicate, think critically, and solve problems creatively.

## References

Armstrong, Thomas. (2006). *The best schools: How human development research should inform educational practice.* Alexandria, VA: Association for Supervision and Curriculum Development.

Berube, Claire. (2004). Are standards preventing good teaching? *The Clearing House*, *77*(6), 264-267.

Ravitch, Diane (2010). *The death and life of the great American school system: How testing and choice are undermining education.* New York, NY: Basic Books.




***Paperback: 320 pages***
***Publisher: Wood N Barnes (June 1, 2014)***
***ISBN-10: 1939019125***
***ISBN-13: 978-1939019127***

## *Connecting Children to Nature*

***By Michael Bentley, Michael Mueller, & Bruce Martin***

***Review by Anne Mannarino***

I have just finished reading the book: *Connecting Children to Nature* by Michael Bentley, Michael Mueller, and Bruce Martin and realized how often we dismiss the outdoors as a learning environment. As a science educator, and also informal environmental workshop planner, I found the book both informative and engaging on all levels. The book is divided into 5 sections that explores why children need to investigate nature, teaches how to develop empathy for nature, shows various ways to explore the natural environment, identifies how children can get actively involved in social action through citizen science, and finally lists multiple resources. The book tells many personal stories from multiple authors and also presents a historical perspective on the nature of science. It is not just a "how to" book as it cites solid research that support how you can make a successful plan to connect children to nature. You can read the entire book or parts of this book and come away with ideas, proven activities, and resources to guide you into action to start connecting your students or your own children to nature. I enjoyed the book so much that I am now ready to plan my next outdoors group meditation (p.134). If you are planning any outdoor environmental activities or programs, this book is an excellent resource as well as a planning tool for educators, parents, and any nature lover.