The Impact of Student Literacy on the Living Environment Regents Exam

Scott Partridge
St. John Fisher College

Recommended Citation

Please note that the Recommended Citation provides general citation information and may not be appropriate for your discipline. To receive help in creating a citation based on your discipline, please visit http://libguides.sjfc.edu/citations.

How has open access to Fisher Digital Publications benefited you?
Follow this and additional works at: http://fisherpub.sjfc.edu/mathcs_etd_masters

This document is posted at http://fisherpub.sjfc.edu/mathcs_etd_masters/33 and is brought to you for free and open access by Fisher Digital Publications at St. John Fisher College. For more information, please contact fisherpub@sjfc.edu.
Abstract
This research study was conducted to determine the affect of reading ability on the success rates of 8th grade science students that take the New York State Regents Living Environment Exam. There is a concern that the English/Language Arts skills of current 8th grade students, not content knowledge, will limit the students' ability to be successful on the Regents level exam if it is given in 8th grade to all science students. This study administered two versions of a past Regents exam to six groups of 8th grade science students. one version of the test that was given was composed of thirty multiple choice questions worded as they appeared exactly on the test. The second version was composed of the same questions that were re-worded to be more "ELA friendly" for the students taking that test. The data strongly supports that students will have a lower success rate when taking tests with a higher reading level. Re-wording difficult tests can improve student achievement without changing the content that is being tested. Students that take the Regent Living Environment course may still struggle on the exam if their ELA skills are not as equally strong as their biology content knowledge and skills.

Document Type
Thesis

Degree Name
MS in Mathematics, Science, and Technology Education

This thesis is available at Fisher Digital Publications: http://fisherpub.sjfc.edu/mathcs_etd_masters/33
1-1-2007

The Impact of Student Literacy on the Living Environment Regents Exam

Scott Partridge
St. John Fisher College

Follow this and additional works at: http://fisherpub.sjfc.edu/mathcs_etd_masters

Recommended Citation
The Impact of Student Literacy on the Living Environment Regents Exam

Abstract
This research study was conducted to determine the affect of reading ability on the success rates of 8th grade science students that take the New York State Regents Living Environment Exam. There is a concern that the English/Language Arts skills of current 8th grade students, not content knowledge, will limit the students' ability to be successful on the Regents level exam if it is given in 8th grade to all science students. This study administered two versions of a past Regents exam to six groups of 8th grade science students. one version of the test that was given was composed of thirty multiple choice questions worded as they appeared exactly on the test. The second version was composed of the same questions that were re-worded to be more "ELA friendly" for the students taking that test. The data strongly supports that students will have a lower success rate when taking tests with a higher reading level. Re-wording difficult tests can improve student achievement without changing the content that is being tested. Students that take the Regent Living Environment course may still struggle on the exam if their ELA skills are not as equally strong as their biology content knowledge and skills.

Document Type
Thesis

Degree Name
MS in Mathematics, Science, and Technology Education
The Impact of Student Literacy on the Living Environment Regents Exam

Scott Partridge

St. John Fisher College
Abstract

This research study was conducted to determine the affect of reading ability on the success rates of 8th grade science students that take the New York State Regents Living Environment Exam. There is a concern that the English/Language Arts skills of current 8th grade students, not content knowledge, will limit the students' ability to be successful on the Regents level exam if it is given in 8th grade to all science students. This study administered two versions of a past Regents exam to six groups of 8th grade science students. One version of the test that was given was composed of thirty multiple choice questions worded as they appeared exactly on the test. The second version was composed of the same questions that were re-worded to be more "ELA friendly" for the students taking that test. The data strongly supports that students will have a lower success rate when taking tests with a higher reading level. Re-wording difficult tests can improve student achievement without changing the content that is being tested. Students that take the Regent Living Environment course may still struggle on the exam if their ELA skills are not as equally strong as their biology content knowledge and skills.
# Table of Contents

Abstract 2

Introduction 4

Literature Review 6

*Readability* 7

*Student metacognition and reading skills* 11

*Test accommodations* 17

*Teachers developing practice* 20

*Gender difference* 25

*Summary* 26

Methodology 27

*Participants* 27

*Instruments and Materials* 29

*Data Collection* 30

*Procedures* 31

Results and Analysis 32

Table 1 33

Discussion and Conclusion 36

*Interpretation and Insights* 36

*Recommendation for Future Research* 42

References 45

Appendix A: Test Version A 48

Appendix B: Test Version B 53
The Impact of Student Literacy on the Living Environment Regents Exam

In the near future we might begin to see that some school districts make the move to having all students take the New York State Regents Living Environment Exam during their eighth grade school year. It is currently a common practice to have eighth grade science students, who are capable of being accelerated and enrolled in a Regents level courses, enrolled in the Living Environment course. All of the students will be assessed at the end of the year by taking the New York State Regents Living Environment Exam, which is a very rigorous and intense test that uses all forms of questioning from multiple choice and matching to constructed response.

As the accelerated science and general science programs continue to excel at extremely high rates of mastery at the eighth grade level, the question is raised as to why teachers do not accelerate all of our students rather than just a small percentage? This question is the driving force behind a proposal to have all students in the eighth grade complete the Regents Living Environment course. There is one major concern with this proposal: are the students' literacy, language arts, and reading comprehension skills strong enough to allow the students to be successful in this course? This question then leads to additional questions that further directed our research. What has a greater impact on the students' success rate on the Regents exam, their literacy skills, content knowledge, or both? It was our belief that difficulty of vocabulary and reading levels of the questions had a greater impact on student success than the actual concepts and understandings taught throughout the year.

As these concerns were discussed, the researchers decided that they need to answer the questions that have been raised, and then come up with a solution. How do
they determine if the literacy is the greatest limiting factor for the students? One possible method would be to see how re-wording the questions, making them more reader or literacy-friendly, impacted the student success rate. Does re-wording Regents test questions increase student success without invalidating the assessment itself? If the students' success rate increased significantly on the literacy-friendly test, then it could concluded that students' success is greatly impacted by literacy and teachers must find a way to help students overcome this obstacle themselves before and during the exam.
Literature Review

There are numerous research ideas and debates being thrown around the education field about how to determine, assess, and improve the literacy levels that students must obtain to be successful on classroom assessments in all areas of content. Educators and researchers first needed to establish the difference between readability and literacy before they could further rate the levels and effectiveness of given assessments for these two such terms (Fry, 2002; Roberts & Gott, 2004). There are numerous types of formulas that claim to measure readability and literacy (Homan, Hewitt, Linder, 1994; Meyer, 2003). Measurements of readability have been compared to determine how changes in readability impacts both student performance and test validity (Hewitt & Homan, 2004; Paul, Nibbelink, Hoover, 1986; Shorrock-Taylor & Hargreaves, 2000; Wiggins, 1998). Student metacognition and testing skills have been influenced by readability and literacy, forcing teachers to struggle with student differences, modeling, communication barriers, and ultimately success of their students (Barba, Pang, Santa Cruz, 1993; Bhattacharya, 2006; Fordham, 2006; Leal, Johanson, Toth, 2004; Lee & Fradd, 1998; Rickey & Stacy, 2000). Such differences have caused many debates about how much accommodating can be done when writing tests to change test question length, difficulty in vocabulary, and even teacher guidance provided to the students without decreasing that tests validity (Gunning, 2003; McCallister, 1930; Wiggins, 1998).

Research had also provided valuable strategies and ideas to assist educators to overcome student attitudes, and improve their own professional practice as a teacher that to improve success of all boys and girls (Branscomb, 1981; Digisi & Yore, 1992;
McCallister, 1930; Tamir, 1988) despite potential differences in science performance between genders for various subject matters.

Readability

Before educators and creators of assessments can debate the level of readability and literacy on assessments or tests that they plan to work with, they first must clarify what is meant by the terms literacy and readability. Roberts and Gott (2004) stated that literacy, when used in discussing scientific literacy, is not so much related to reading levels but rather “understanding of how science has developed and its contribution to society” (p. 9). This use of literacy was more focused on the “questioning of evidence” (Roberts & Gott, 2004) in science and not so much the reading level in science text (p. 9). Which caused this research to focus its’ attention to the term readability. “Readability is defined as the ease of comprehension of style of writing” (Fry, 2002, p. 286). This definition of readability is often confused with the term leveling which was defined as “selecting books to match the competencies of the reader” (Fry, 2002, p. 286). There seems to be very little difference between the two terms when they are used quickly in conversation about selection of books for students to read at various levels. However, further exploration into each term created a much larger gap between the true meaning and uses of each form of measure. Fry (2002) claimed that the traditional readability formulas that have been used to determine given difficulties for texts are based on two major criteria: syntactic difficulty and semantic difficulty. Syntactic difficulty mainly is the analysis of sentence length, how many words are in each sentence? This can also be referred to as grammatical complexity. Semantic difficulty is not the measure of sentence length but rather the measure of each word. This answered questions like how many
syllables are there. How often is that word used? Is this word something that is familiar to the reader? These two forms of measure, syntactic and semantic difficulty, allows for readability scores that are determined to be very objective. “Simply type in a passage or scan in a whole book and the computer will give you a readability formula score” (Fry, 2002, p. 287). This raised the concern about the effectiveness of using readability in a field like education where no one child is the same.

Fry (2002) lead readers to thinking about how using leveling in education seems to make much more sense since educators have to adjust and accommodate for all levels and skills of students. Leveling takes into consideration content, illustrations, length, curriculum, language structure, format, and judgment as a means to determine level of difficulty for a given text. All of these criteria make it difficult to apply any one objective score to a text. Lack of such objectiveness causes the measure of leveling to be used strictly within the classroom; where as readability was and is applied to all areas within a classroom and in the real world. As researchers try to determine the level and impact of readability and literacy on science assessment scores, there needs to be clarity as to which term is truly has been measured; literacy, readability, or leveling. As the research is focused on reading difficulty of text on assessments the terms readability and leveling should be considered; where literacy levels is focused around the understanding of how science works.

Selecting text that has appropriate readability and leveling for students is essential, and the research showed that there were many flaws in how developers have used and applied readability formulas and scores. Lack of effective control on readability can create major issues; most importantly the knowledge of what caused the incorrect
responses from students. Incorrect answers might have been from the "lack of knowledge or because of the inability to successfully comprehend the test and its items" (Homan, Hewitt, & Linder, 1994, p.349). Teachers do not want "packaging of the text to interfere with the student’s understanding" (Meyer, 2003, p. 204) or ability to demonstrate their true understanding. At first glance, research completed by Meyer (2003) showed that "readability formulas provide an easy and quick way to predict readability" and that there are over 30 different formulas that can be used to determine readability of text (p.206). The research was then quick to state how there were no strict guidelines in place for writing text at given levels for readability. Meyer (2003) pointed out that "classic readability formulas focus on word and sentence characteristics" (p. 207) making measured readability on assessments very quantitative. These easy formulas lacked means to measure important qualitative characteristics like student judgments, reactions, and overall comprehension of content. This creates text that lacks coherence which focuses on the "overall logical structure of the text, the top-level structure, as well as cohesion, how well the parts of the text hold together" and "coherent, well-organized texts are easier to understand" (Meyer, 2003, p. 207).

Even if readings are more coherent, does it always mean that student success will be higher? Research completed by Paul, Nibbelink, and Hoover (1986) was completed as a means to discover if lower readability scores also lead to higher success rates despite the level of content mastery required. Fifteen selected math questions were each put into verbal form at three different levels of readability and it resulted in six forms of the math literacy test. Results indicated that "whether a story or problem has a readability score a few grades below, at, or above grade level, there is no substantive effect on students
ability to solve it” (Paul, Nibbelink, & Hoover, 1986, p.170) and resulted in the conclusion that readability did not have a major impact on success, rather it was mostly the knowledge of the content. This was contradicted by the research completed by Hewitt and Homan (2004) that explored how the major shift to using standardized tests as the major means of assess caused the need for readability scores to be valid. The Homan-Hewitt Readability Formula examines individual item readability across three grade levels with the belief that “the higher the item readability, the more the students miss that item” (Hewitt & Homan, 2004, p.1). Scoring and comparison of the standardized tests had added value because of the major shift toward schools relying strictly on standardized tests for student and program assessment. The research showed occasional instances where a high level readability question had an attractive right answer, or a low readability score was about content not covered in class, leading students getting the questions right or wrong for reasons other then readability. The results also allowed Hewitt and Homan to be able to safely state that “readability level is often a major factor” in students’ success but caution others to know that it is not the only factor (p.14). These results varied from previous research by Homan and Hewitt along with Linder that had stated “As readability level increased, item difficulty was affected as predicted” (Homan, Hewitt, & Linder, 1994, p. 356). They had used the Homan-Hewitt Readability Formula on this assessment that consisted of 12-questions at seven different levels, and as readability increased or decreased performance followed it closely.

The differences in results varying from readability having no impact, readability has some impact, and readability having great impact on text difficulty cause for there to be a greater amount of research completed to draw stronger conclusions about the impact
of readability on assessment performance. “However, reverting to the main focus of this study, the research is needed above all to support the best endeavors of test developers in enhancing the validity of the assessments” (Shorrocks-Taylor & Hargreaves, 2000, pp. 59-60). If more research was needed to improve assessment validity, someone can raise the question of why do we need to improve validity? Grant Wiggins (1998) made sure that there is a known difference between a test being authentic, and valid. Authentic tests can be meaningful, but do they test what they are supposed to be testing? Wiggins (1998) defines tests as being valid by stating:

Tests are valid if we can infer real performance results for specific standards from test results. Many authentic tasks make this difficult because there are so many simultaneous variables at work. Maybe the student’s excellent performance was due more to native intelligence than to mastery of the content-control over the latter being the reason the task was designed. (p.141)

If readability has a great influence on text difficulty, then too high of readability or too low of readability of test questions for certain levels of test can easily cause a test to simply be testing reading ability of the student rather then understanding and mastery of content knowledge. If a test is meant to test content, and it truly tests readability then that test would be considered invalid. More research has to be completed to determine effectiveness of readability and its impact on test validity.

Student metacognition and reading skills

There is an unlimited amount of research available that supports the idea that all students are different, and each student will have different strengths and weaknesses about what they like to do and can do in the classroom, the real world, and in their minds.
The skills required for students to be successful at performance in all areas differ for each individual. The challenge to educators and assessment creators is being able to adjust to all the metacognitive differences there are in every classroom environment.

Metacognition is described as being “knowledge concerning one’s own cognitive processes and products or anything related to them” (Rickey & Stacy, 2000, p. 915). In other words, metacognition is someone’s ability to think about how they think. For students and teachers to know and understand about how someone thinks is a great asset since research has shown that metacognition greatly influences deeper learning for two reasons: “awareness of one’s own thoughts is important for development of understanding of ideas, and awareness and control of thinking have shown to have significant impact on problem-solving success” (Rickey & Stacy, 2000, p. 915). In classrooms, students with high levels of metacognition are more able to successfully work through difficult understandings on their own rather then someone telling them how to, which leads to a greater and deeper mastery of the knowledge. This success can then allow students to overcome difficult questions, identify naïve results in chemistry experiments, and still be successful despite a possible lack of experience or knowledge. Simply strong metacognitive thinkers can overcompensate for lack on knowledge and still find success on tests. This allows for a classroom of students who encounter a question about a topic not covered by the teacher through the course of the year to have a good chance of still thinking through a problem successfully on Regents exam.

As with vast differences in students ability to read text correctly and think through problems correctly, there is also a great gap in how successful students are at communicating what they are reading and thinking successfully. A major factor in
student success with science literacy is the ability to handle the science language. Many standardized tests and teacher lessons present content in scientifically appropriate ways with language that is used by scientists and science teachers. There is little consideration for the students’ natural literacy and language. Teachers need to find a shared language usage where students and teachers can interact without a focus on specific subject contents. There are questions that teachers have to ask themselves in order to find this balance, like who are my students? What are my students ELA skills? What does the science we are learning require? How can I guide my students to understand this knowledge and language? Answering these questions can promote a dynamic balance that supports both sides of the issue in developing students’ science literacy; what students need and what science demands. Students have to be able to do more the read and write in science, science education also requires “learning to observe, predict, analyze, summarize, and present information in a variety of formats, such as orally, in writing and drawing, and through tables and graphs” (Lee & Fradd, 1998, p.14). Through more developed communication we can make more reliable assessments about the students’ ability to answer questions using logic, reasoning, and critical thinking. Developing and using these skills in a science classroom through use of a shared language can essential for literacy and learning in science and life. Students can get a multiple choice question right whether they have a true understanding of the content or with a random guess, demonstrating science literacy is a two part process that goes beyond just knowing what the correct answer is, but also being able to communicate it, and practice it in the real world. “The development of science knowledge involves
"knowing" science (i.e., scientific understanding), "doing" science (i.e., scientific inquiry), and "talking" science" (Lee & Fradd, 1998, p.15).

Being able to do science in the classroom, completing homework, and even completing labs can greatly be influenced by the ability of the students to simply read the text. A big issue arising in classrooms is the lack of student ability to simply handle the English language itself. How can students be expected to master science content by using the English language when they can’t use English? The numbers of minority groups whose native language is different from English is increasing, and over the years there has been an increased use of reader friendly textbooks that help assist with this issue. Still today most textbooks are not reader friendly to the group that is using them. Often textbooks are not written at low enough levels for certain content areas and the reading ability of most students continues to decrease causing this issue to be getting greater and greater. However, formats are changing to help account for this struggle. “Such conventions as headings and subheadings, glossaries, and visual strategies (such as photographs, diagrams, tables, graphs, and flow charts) are instructional devices within the text that support and illustrate textual material increasing students’ learning” (Barba, Pang, & Santa Cruz, 1993, p.17). As the textbooks shift more and more towards such formats that are friendlier to lower level reading students, there is still the question of how assessments are and can be adjusted to meet the same accommodations?

For students to accurately comprehend information from a reading, they have to be able to read the words accurately and fluently. If there is a struggle with comprehension strictly because of ability to read words, then a gap will be created between students in their ability to understand science content. “It is the ability to read
words accurately and instantaneously that distinguishes children with and without reading difficulties" (Bhattacharya, 2006, p.116). Teachers need to find strategies that can work to help students read words quickly on science test so the students are not getting caught in the difficulty of the words which can be distracting from the content of the question. Polysyllabic words cause a great deal of stress to struggling readers. These words are longer words that have three or more syllables in them, and may often be known to the students by sound but not by sight as they can not read through them. Teachers of students with reading difficulties have to provide a great deal of practice for the students to work through syllables of words to produce a comfort level in the students that can allow them to make it through tests such as the Intermediate-Level Science Examination and the New York State Living Environment Regents Examination. Many of the questions and answers on such exams will be loaded with polysyllabic words like “hereditary, reproduction, photosynthesis, metamorphosis, regulation, fertilization, respiration, circulatory, cementation, sedimentation, crystallization, evaporation, precipitation, condensation, and phenomenon” (Bhattacharya, 2006, p.119) which all have to be read before a student can begin to think about selecting an answer. Sample questions can easily demonstrate how much polysyllabic words appear through such exams.

Which two processes result in the formation of igneous rock?

1. melting and solidification
2. sedimentation and evaporation
3. crystallization and cementation
4. compression and precipitation (University of the State of New York Intermediate-Level Science Examination, 2000, Part A, p.6)
This single question required students to read 10 polysyllabic words out of the 22 before they could even attempt to select the correct answer. This is only one question, and students would be overwhelmed with question after question of the same format of words on a single test. The assessment would be at risk of being invalid based on the fact that incorrect responses may be the result of difficulty with lengthy words rather than the lack of sufficient content knowledge.

Research has been done to determine how much student success could be improved through the use of tutors, and their ability to work with students on their reading experience and reading interest. With many aspects of life, people are willing to put forth more effort and dedication to things that they enjoy and are interested in doing. This can directly be applied to students and learning. Lack of interest can lead to lack of motivation and effort resulting in lack of success on the part of the child. There are three common things that apply to successful tutoring programs: “(a) many opportunities to read authentic materials, (b) many applications of reading integrated with authentic writing experiences, and (c) highly motivated reading and writing activities related to students’ interests and abilities by caring tutors” (Leal, Johansson, & Toth, 2004, p. 76).

Such practices need to be used beyond just tutoring programs; applying these three practices to everyday class experiences can improve student interest in reading as well as success. “Many times if a strategy or activity was not successful, the tutor reported it was due to his or her lack of experience rather than the lack of interest in the child” (Leal, Johansson, & Toth, 2004, p. 81). This data gives support to the idea of making such activities a common practice not only for the students, but also the teachers. Student success is greatly influenced by the confidence and interest of the students who are
learning and the teachers who are teaching. Increasing practice and remediation in reading by students and teachers fosters success for both sides involved, "reading practice- just reading- is a powerful contributor" (Leal, Johansson, & Toth, 2004, p.85) to continued success of increasing reading ability in science.

Test accommodations

As teachers strive to find appropriate reading materials for their students, they must caution themselves about going too far. Just as it is important to not select reading that is too difficult for students and causing them to fail, teachers have to make sure they don’t select any reading that is too easy for students. If reading level is too low, students will not be challenged and then growth of the individuals is then not maximized. Throughout the years, many factors are taken into consideration by teachers as they have selected reading material at proper levels for their students, but in the end "the two factors that most efficiently predict the difficulty of a text turned out to be vocabulary and sentence length" (Gunning, 2003, p. 175). With other factors kept the same (like student ability, content knowledge, and reading environment) it is more difficult for students to accurately read sentences that are longer. Also, with other factors kept the same, sentences that have difficult vocabulary are harder for students to read then sentences with simpler words. As tests are written, research has found that most tests experienced by students at certain grade levels are written at or above that grade level. This presents a larger problem when most students at given grade levels are actually below reading ability for that level which widens the gap of the questioning level and student ability even more. This brings about an even greater need to establish and implement more readability and leveling measures in our classrooms. Teachers must find more guidance
in finding appropriate levels of reading, and put these measures into practice as they plan for instruction. That way their instruction can show more correlation to the reading levels encountered on the standardized tests that they will encounter. More research is also needed to determine how much adjustment can be made to assessment questions without risking test validity.

As teachers administer tests, they often encounter situations where they need to address students’ questions and provide proper guidance without giving too much assistance and ultimately giving away the answers. But before research can address how much guidance can be given to students, the research needed to address what types of guidance students will need. Students often find themselves needing guidance in various methods used to work through activities and problems, guidance is needed for identification of relationships in reading and thinking that is required. Students also need guidance in how to review material and overcome vocabulary, as well as making accurate interpretation of data. These issues are well known to teachers; however teachers find themselves “at a loss to know how to inject training in reading into courses already crowded with other things” (McCallister, 1930, p. 271) from the first day of class to the last. Research does make it clear that individual and group practice of guidance procedures in everyday class routines greatly increases student success and confidence. Research fails to address not only the time issue allowed for providing proper practice and guidance, but also how much guidance provided to students while testing is too much? Courses require students find solutions on their own, at what point can guidance make it not ‘their own’?
If a test still is able to test what it is supposed to test even when adjustments have been added to meet student needs, then this test still is valid. Wiggins (1998) states that tests are valid if “we can infer real performance results for specific standards from test results” (p.141). Do the results tell teachers what they expected the test to tell them about students’ true understanding? Students with disabilities do need certain accommodations that help with a disability not related to the content knowledge being tested for. This allows for students to show what they know about content despite disabilities. Teachers have to make sure that each accommodation “eliminates at least one source of variance that is not fundamental to the underlying construct” (Fletcher, Francis, Boudousquie, Copeland, Young, Kalinowski, & Vaughn, 2006, p.138) for students. Variances due to disabilities are irrelevant to what the test is expected to measure, and a “valid accommodation will improve performance only for students with a disability” (Fletcher et al., p.138). Teachers need to get proper training to make sure that as tests are created, the tests allows for accommodations while considering the relation of the accommodation to the content knowledge.

There are certain requirements that are upheld when tests are written and these requirements for test must be followed while accommodations are being made for students disabilities and reading difficulties. The Guide for Item Writers: Intermediate Level Science Examination (Parts A, B, and C) lists essential parts of multiple-choice, constructed-response, and extended-response items. Multiple-choice questions are required to include a stem or the point of the item with a set of four alternatives or choices. Three of the four choices must be incorrect, while the last response is correct. Constructed-response items must include a stimulus and central theme and several
independent items based on the stimulus and scoring guidelines. Extended-response items require a set of directions and a scoring guideline that is based on the goal of the item. Model answers are optional to be included with the scoring guidelines. All three of these items have checklists that must be followed to ensure questions are fair. These checklists are on p.13, 18, and 24 in the guide from The Education Department in Albany, NY (1995).

*Teachers developing practice*

In many subject areas (especially math, science, and social studies) "the conceptual load is substantial and associated vocabulary in course texts is technical and intimidating" (Fordham, 2006, p.390). This level of intimidation causes for a great need for teachers to be like coaches to their students. Not only are teachers responsible for the delivery of material and content, but also skills the students needed to be successful and confident in their studies. Many teachers in the science field originally did not begin their careers as educators; rather they were just experts of science knowledge and not working with children. Lacking the expertise of education of youth can cause for teachers to be misled by student achievement, not only during instruction but also on assessments. Teachers may “mistakenly feel that their assessment-oriented questions are sufficient to help struggling comprehension” (Fordham, 2006, p.394). Professional development of teachers practice needs to help teachers to become more aware of their practice, teachers need to be acting more as guides then assigners in their content. Professional development can be in many forms, starting with observation of practice and building to continued practice of not only assessing practice but also addressing it.
Research has also brought to light the issue that science teachers need to be prepared to address issues students might have with religious and philosophical beliefs when being compared to science concepts and theories. Students will bring various bias and humanist ideas that may surface at inappropriate times in a lesson. Teachers who are unprepared may not handle such issues might get caught in the trap of debating and arguing rather than discussing issues of traditional practice and beliefs. Educators need to act as coaches to help deliver science facts and “assure that such postulations are clearly articulated to the users” (Branscomb, 1981, p.9) whether talking with religious leaders or young teenage minds. Lack of such sensitivity toward fiction can cause student attitudes to become negative towards the content and interest can be affected.

There are certain methods of professional development that are more effective than others in helping teachers to become more proficient in content delivery to increase student success. Teachers need to be given additional practice rather than simply being told what to do. Teachers also will value their learning experiences even more when they play a role in the research, collaborating with the researchers to jointly address the issues. Working with a peer as a support network can help add different points of view and create dialogue around various issues and dilemmas. “Ongoing opportunities for collaborative professional development” (Pedrotty-Bryant, Vaughn, Linan-Thompson, Ugel, Harnff, & Hougen, 2000, p.239) will allow for continued adjustment in professional practice which will allow for teachers to recognize and adjust to their practice as needs of the students changes. Research has gathered data about three types of reading strategies (collaborative strategic, partner reading, and word identification reading strategies) and influences on classroom success. Students made gains in all three
Literacy on Regents

areas when used in practice within the classroom leading to greater student success. Teachers not only were able to see how the children benefited, but as professionals who received proper professional development they greatly benefited in their own ability to working with reading difficulties.

Whether teachers receive professional development beyond their classroom or not, the fact still remains that they will be responsible for their students’ performance on written tests. This responsibility has often led to a teacher who teaches to the test. This practice is one that makes sense when looked at from the teacher’s point of view. Teaching to the tests will focus the classroom activities, time, and effort of the teacher around the needed skills required to complete that test. Students will have a increased chance of being successful on that given test, scores will be higher, and goals will be achieved from the teachers point of view. But what does that mean for the students? “Would teachers just ‘teach to the test’, resulting in the less practical work in the classroom?” (Roberts & Gott, 2004, p.20). Our overall aim in teaching is to go beyond basic knowledge of individual concepts and paper and pencil assessments that test for such concepts. Many written tests “form a necessary but insufficient condition for both practical problem solving and an empowering and critical form of science literacy” (Roberts & Gott, 2004, p.19). What we are testing for with many standardized written tests in science is the “understanding and application of procedural ideas” that are somewhat needed for problem solving, but this understanding and application doesn’t constitute the “synthesis of both substantive and procedural ideas in the solutions of the problems” (Roberts & Gott, 2004, p.19). Written tests are able to let us to know if students can read, speak, and know science, but these tests are limited in how much they
can inform us about students’ ability to do science and apply science literacy. As professionals, teachers need to learn how to create more assessments that test the ability to do science rather than recall knowledge. Science exams need to be created to have questions that “allow flexibility for students in their answers; and, most importantly, they should test a wider range of skills than the mere recall of facts” (Roberts & Gott, 2004, p.19). Both sides of the table have to make this change for there to be success, if it is a one sided change the growth of students ability will not be able to be assessed properly. Teachers practice must reflect more “doing science” and less memorizing knowledge. Assessments need to reflect more synthesis of knowledge and less recall.

Creation of appropriate assessments for science literacy has to take into consideration two factors as teachers move to create assessments that increase higher levels of thinking. First, teachers must further research “what type of metacognitive instruction is appropriate for science students at each grade level, reading level, and level of cognitive ability” (DiGisi & Yore, 1992, p.10). Second, teachers must step back and determine “how much instructional time and practice with metacognitive reading strategies is needed for students at each grade level” (DiGisi & Yore, p.10). Training and practice of teachers must be provided so each individual teacher can assess their students and their classroom lessons to meet the needs that may be required for each individual class and each individual students. There is continued change in ability of students from year to year, and proper training needs to be provided for teachers to efficiently adjust their teaching to meet these needs. Do teachers have the ability to instruct students about metacognitive abilities and strategies? Surveys have shown that most secondary science teachers “value the importance of reading instruction in science, and have a positive
attitude about enrolling in content reading course” (DiGisi & Yore, p.11). But these teachers also recognize “their lack of knowledge about how to integrate reading into their instruction” (DiGisi & Yore, p.11) efficiently without taking away from instruction time that is already pressed to cover the curriculum guidelines that will be assessed on state exams. Most professional development that is provided to teachers is a “quick overview of reading and speak about general strategies for helping students to read in content subjects,” (DiGisi & Yore, p. 12). Teachers need to seek out detailed training or courses that are “specifically tailored to science teachers’ needs” (DiGisi & Yore, p.12) that provide strategies that are proven to be efficient in science, rather than courses that cover all types of strategies used in all subjects areas for reading content. Where ever teachers turn to in order to improve their ability to teach content reading in their science classrooms, if they start with the mindset that their students need to “effectively read and learn from scientific text” (DiGisi & Yore, p.15) they will increase the odds that their students will leave the classroom knowing the science content they were taught.

Teachers who believe that “both good and poor readers will profit from such guidance” (McCallister, 1930, p.200) will find that they have good readers who increase their own strategies. They also will have poor readers who will gain new skills and assistance in overcoming their disabilities. “When reading difficulties arise in the study of a subject, it is not unreasonable to expect an instructor to accept the responsibility to aid their pupils to overcome these difficulties” (McCallister, p. 201). As time continues, reading guidance from teachers should become a function of everyday instruction and all students should be able to apply this knowledge and skill to state exams so they can demonstrate
mastery in science literacy. This will ensure that the test is valid and does test students understanding and is not testing reading ability.

**Gender difference**

As research has been done about science literacy, readability, reading skills, and teacher practice, there is one other potential difference among students that could be a factor and impact student success on state exams. Research has shown that there is a gender gap between males and females and their performance in science content. Tamir (1988) discussed how British researchers argue that boys tend to outperform girls on multiple choice items on science tests (Harding, 1983; Johnson et al., 1983). The Second International Science Study (SISS) was performed in 26 different countries during the years from 1983-1984. The study was to assess overall student success for students at the ages of 10 years old, 15 years old and 17 years old. This range of ages can show direct relation to that of students who will be asked to take the Regents Living Environment Exam. The results of the study showed that “boys excelled in Application in all the sciences except biology” (Tamir, 1988, p.143). “Although overall boys and girls who major in biology achieve equally well there are certain topics in which girls excel in Application, especially in topics related to human biology” (Tamir, p.134). Teachers of science should find strategies that will help to overcome such differences in ability and interests that result from differences in gender. Teachers who are aware that certain topics may not interest the male or female students as much can then adjust their practice to overcome such differences just like they should for reading ability, and science literacy skills. Proper adjustments might overcome the belief that “girls in general are less interested in studying science, have less positive attitudes toward science and toward
science learning, have less understanding of the nature of science and are less inclined to continue to study science” (Tamir, p.137). But the research also clarified that in the area of biology and even more so in human biology, this statement does not apply or even might be reversed since there does not tend to be a difference according to gender.

Summary

As educators adjust their practice in science classrooms to address differences in students reading ability, metacognitive skills, science literacy skills, skills for guidance in reading content, and ultimately strategies for improving student success, they also need to consider (except in the area of biology) that gender differences and attitudes can affect the students’ ability to master and demonstrate science literacy skills. Educators that learn to maximize their practice and provide the highest quality instruction as possible should be able to help students overcome vast differences that they share within a single classroom. Overcoming weaknesses in literacy, reading comprehension, and metacognitive skills would allow students to successfully overcome the challenges of they faced on standardized tests, preventing weaknesses in ELA skills from disrupting communication of science knowledge.
Methodology

The following section describes how the study was conducted, who the participants were, the tests that were given, and how the results were gathered.

Participants

Because the Wayne Central School District desires more of its 8th-grade students to take the LE Regents, preferably starting in the 2007-2008 school year, participants for the study were selected from 8th-grade science students enrolled 2006-2007 at Wayne Central Middle School. Although three teachers at the school teach science to 8th-grade students, only student work from two of the three teachers was studied. One participating teacher taught LE Regents to two classes of accelerated 8th-grade students; the other teacher taught physical science to four sections of standard-track 8th-grade students. A total of 40 accelerated students (20 of them female) and 83 regular Physical Science students (43 of them female) participated in the study. No significant difference in abilities was noted among the four classes of regular 8th-grade students in the school that were not selected for the study (taught by the non-participating teacher); logistics and student schedule were the mere determining factors as to which standard-tracked students’ work was studied. However, practice tests are regularly administered to all 8th-grade science students in this district in preparation for midterm and NYS Grade-8 exams, and all 8th-graders received comparable practice tests regardless of their participation in this study.

The classroom environment for the accelerated students is relatively fast-paced, in-depth, and inquiry-based. These apparently self-motivated students demonstrated far-above-average science skills, as well as language skills, personal responsibility, and
overall maturity by the end of 6th grade, which allowed them entrance into the accelerated science program. Participation in the accelerated program was voluntary and students needed to maintain a grade point average of at least an 85% in the class during 7th grade to remain. These 8th-grade students studied earth science with all students in 6th grade, physical science in 7th grade solely with their accelerated peers, and studied Regents-level LE in the 2006-2007 school year. Past classes with similar demographics taught by this teacher have attained 100% mastery of the high-school level material, as measured by a score of 85% or higher on the LE Regents exam. However, it should be noted that these students had not had formal training in the life sciences before this year; it was not expected that they would know all of the material on this study’s test (administered to them in January) because the Regents course continues on through June.

The students in the regular science track demonstrated a wider range of abilities than those in the accelerated program. Because of the less homogeneous nature of the regularly-tracked students, the classroom environment for the regular students was slower-paced and more differentiated in nature than the classes with the accelerated students. It was still inquiry-based and placed a similar emphasis on lab work, collaboration, and mastery of given material. Several of the students in each of the four classes maintained A’s or A+’s and had academic characteristics similar to those of the accelerated students. Most, however, fell in the average or below-average range regarding their demonstration of scientific knowledge and abilities. One of the four physical science classes was considered to be the “blended” class and had a special education teacher who pushed in daily. Although this special education teacher was formally a consultant teacher to only two of the students in the room, 10 of the 22
students in the class had some sort of Individualized Education Plan (IEP) or formalized New York State 504 plan to help them overcome their learning disabilities. Thus the special education teacher assisted many of this class's students in various ways. Most of these students' IEP's or 504 plans existed to address language-based issues the students had; these students read at a grade level significantly below the average 8th grader. The other three physical science classes used in this study had a total of seven (of the 61 total) students with an IEP or 504 plan, all of them due at least in part to language-based reasons. One of these seven students demonstrated particularly strong science skills; the student's A+ average was hindered only by the occasional language-based difficulty regarding following directions or understanding what was being asked.

All of the physical science students in the study had studied life science in 7th grade, and earth science in 6th grade. Although the biology required of these students at the 7th-grade level was less than that for the Regents LE level, the district consistently had over 95% of its students pass the New York State 8th grade science assessment, and the 7th-grade curriculum is rather well-aligned with the material that the students need for the LE Regents. The 8th-grade students in physical science classes are expected to take the actual LE Regents course and subsequent exam in their 10-grade year. So, like the accelerated students, the students in the physical science classes have not yet taken all the coursework required to take a real LE Regents exam.

Instruments and Materials

This study utilized the fact that all 8th-grade students at Wayne Central were required to take various practice science tests in preparation for their January midterm exams and their heavily emphasized New York State 8th-grade science assessment. Such
practice tests were commonly used to refresh students' memories of material, practice utilizing wise test-taking strategies, better predict which concepts still need teaching or reteaching, and generally get the students in the best mindset for high-stakes science test taking. Student test scores used for this study did not count in any way towards student grades. Validity of the results might have been somewhat compromised since students might not have done their best on a practice exam, but students also knew that any poor practice test scores could result in their being on a special after-school help schedule, so some motivation was there for students to take this test seriously.

Students in one of the two accelerated classes and students in two of the four regular physical science classes were given 30 multiple-choice questions from the actual LE Regents exam. This test was termed Version A (Appendix A). The students in the other three classes received a language-friendlier version, Version B (Appendix B) of the same exam. Science vocabulary and concepts were unaltered in order to best isolate readability and language issues as the manipulated variable; sophisticated non-science vocabulary, sentence structure, word density, question and response length, etc., were the only aspects of the exam that were changed. Neither group of students was told that the test questions were Regents-level, but students were warned that they had not yet been exposed to some of the test material. The teachers merely told the students to "do the best you can" and "use your test strategies" if students mentioned that they "don't know or remember any of this stuff."

Data Collection

Student test scores from both Version A and Version B were compared, as were the accelerated students' scores versus the physical science students' scores. Specific
questions, correct responses, and detractors were analyzed for readability and language-based issues versus science conceptual or vocabulary factors. Test latency (the time students needed to complete the exam) was noted. All students were given about 80 minutes to complete the 30 multiple-choice questions to help to assure that the students in all six classes had more than ample time to perform their best on the exam.

Procedures

Standard Regents test-taking procedures were utilized. Students used a #2-sized pencil and received both a test booklet and a corresponding Scan-Tron answer sheet on which they were required to bubble in their answer to the multiple-choice questions. Student talking or collaboration was not allowed, and students were proctored to help eliminate any cheating. The allotted timeframe for the exam was made known to the students and was posted visibly for all to see. Students raised their hand and waited until a teacher assisted them if they had any questions or concerns. Students were told to use their test-taking strategies, which have been emphasized since 6th grade, to help maximize their achievement. Students with IEP’s or 504 plans did not receive any of their usual testing modifications for this study. Such modifications vary with each individual, but include accommodations such as taking exams in a separate location to minimize distractions, allowing for subvocalization, reading the test questions to the student, checking for understanding, and rephrasing directions.
Results and Analysis

Each of the six groups (A, B, C, D, E, and F) of students were tested with either Version A or Version B. The average number of questions correct out of 30 total questions was calculated for each group. These results were then converted to the average percentage correct for each group. The times it took from the beginning of the test for the first student to finish, 80% of the students to finish, and all the students to finish were recorded in minutes for each group. The researchers also gathered the average midterm achievement grades for each group to be used as a comparison as well. These results can be found in Table 1 on the next page. Students who were given the actual LE Regents Exam, Version A, scored an average of 52% correct. The students who took the re-worded version of the test, Version B, scored an average 61% correct. Results needed to be further broken down to allow the researchers to make a comparison between the accelerated classes and the general classes.

Group A (8th grade accelerated Biology students taking test Version A) consisted of 17 students. The average number correct was 19.1 out of 30, which converted to 64% correct. The first student was finished in 21 minutes, 80% of the students in Group A finished in 30 minutes, and all the students were finished in 43 minutes. The average midterm achievement score for this group was 81.8%.

Group B (8th grade accelerated Biology students taking test Version B) consisted of 23 students. The average number correct was 23.2 out of 30, which converted to 77% correct. The first student was finished in 13 minutes, 80% of the students in Group B finished in 23 minutes, and all the students were finished in 28 minutes. The average midterm achievement score for this group was 82.2%.
Table 1

*Student Performance on Actual and Reworded LE Regents Questions*

<table>
<thead>
<tr>
<th>Student Group</th>
<th>% Correct</th>
<th>Midterm grade</th>
<th>Time for 80% of exams completed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Version A (actual)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accelerated class 1</td>
<td>64</td>
<td>82</td>
<td>30 min.</td>
</tr>
<tr>
<td>Regular class 1</td>
<td>48</td>
<td>83</td>
<td>---</td>
</tr>
<tr>
<td>Regular class 2</td>
<td>44</td>
<td>80</td>
<td>---</td>
</tr>
<tr>
<td>Test Version B (reworded)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accelerated class 2</td>
<td>77</td>
<td>82</td>
<td>23 min.</td>
</tr>
<tr>
<td>Regular class 3</td>
<td>55</td>
<td>75</td>
<td>---</td>
</tr>
<tr>
<td>Regular blended class</td>
<td>50</td>
<td>76</td>
<td>---</td>
</tr>
</tbody>
</table>

*Note.* Dashes indicate test completion time for regular students was not recorded. Accelerated students took a different midterm exam than the regular-track students.
Group C (8th grade general science students taking test Version A) consisted of 21 students. The average number correct was 14.4 out of 30, which converted to 48% correct. The average midterm achievement score for this group was 82.6%.

Group D (8th grade general science students taking test Version A) consisted of 21 students. The average number correct was 13.1 out of 30, which converted to 43.8% correct. The average midterm achievement score for this group was 80.2%.

Group E (8th grade general science students taking test Version B) consisted of 22 students. The average number correct was 15.0 out of 30, which converted to 50.0% correct. The average midterm achievement score for this group was 76.3%.

Group F (8th grade general science students taking test Version B) consisted of 19 students. The average number correct was 16.5 out of 30, which converted to 55.0% correct. The average midterm achievement score for this group was 75.3%.

As expected the accelerated students in Group A and B scored a significantly higher average than the general science students. The midterm scores showed that Group A and Group B had equal ability based on previous achievement scores. This further helps support our hypothesis based on the results. Accelerated students who took the actual test scored lower average scores (64% correct) on the 30 question test than students of equal ability who had taken a re-worded form of the same test and scored higher (77% correct). The four groups of general science students' test results gave even stronger support that reading ability greatly impacts test results. When the four groups are separated by midterm scores, we can group them into two. Groups C and D have similar midterm scores showing they are about equal in ability, both around 80%. Groups E and F midterm scores are very similar and significantly lower, 75% and 76%.
Groups E and F students that performed at a lower level then groups C and D on midterm tests, took the reworded version of the test and scored 50% and 55%. Groups C and D, the two general 8th grade classes performing at a higher level, scored 48% and 44% correct on the test Version A. These numbers showed the researchers that students with lower abilities scored higher because they took a re-worded form of the test. And looking at it from the other point of view, students of higher ability scored lower because of the higher reading level of the test. The research and data strongly supports that students will have a lower success rate when taking tests with a higher reading level. Re-wording difficult tests can improve student achievement without changing the content that is being tested.


Discussion and Conclusion

In the groups of students that were tested, the average test scores were higher across the board for the three groups that took Test Version B. When the test questions and answers were re-worded to make them more ELA friendly students were better able to respond correctly. The increased literacy level required for achieving success on the actual worded test Version A was a direct result of longer sentences, increased difficulty with vocabulary, and overall test readability. Accelerated 8\textsuperscript{th} grade students that took the re-worded test, Version B, achieved a 13\% higher average score than the accelerated students with equal abilities who took the actual version of the test. The general 8\textsuperscript{th} grade classes had similar results that almost mirrored the accelerated scores. The classes that took test Version B scored an average 7\% higher then students who had taken test Version A. The teacher also had noted that the two general 8\textsuperscript{th} grade classes that took the reworded test are blended classes and normally perform around 6\% lower then the other general 8\textsuperscript{th} grade classes that took the actual test. When you account for both those percentage differences, the two classes that took the re-worded test overcame the normal 6\% difference and went above and beyond by 7\%. Now we have a difference between classes that matches the accelerated classes' difference; students of all abilities achieved 13\% more success with a re-worded version of the test.

Interpretation and Insights

The purpose of the NYS Living Environment Regents Exam is to assess the students' knowledge and skills about life science and the living environment. The test should allow the students to demonstrate and express their knowledge and understandings relevant to the content that they have been studying. They validity of this exam can now
be questioned when literacy is looked at. A student's success or failure on this exam should be only determined by the student's ability to demonstrate their content knowledge of the living environment. "Tests are valid if we can infer real performance results for specific standards from test results" (Wiggins, 1998, p. 141). The validity of this exam can now be questioned, where students are being asked to show not just content knowledge, but also English/Language Arts skills. If a student lacks the ability to read and decipher Regents level questions or lacks the ability to express their thoughts in writing, they will stand no chance at success on the exam even if their living environment content knowledge is at mastery. This specific test according to Wiggins (1998) could be considered invalid since it is testing more then content knowledge of the living environment.

The research found that there was a 13% average difference in score on a test that has been reworded. That difference was only an average for students of all abilities. The 13% average difference was present between the classes with the school's top 25% of students who are extremely strong in ELA, and the same 13% difference in success was present in classes of students with average abilities. The results are only showing averages, one can only imagine how big the difference in achievement scores might have been for students who do struggle with ELA. If this test was given to two groups of students who were in the lowest 25% of students in the school in ELA, we can only imagine how vast the average difference would be. This thinking might be a little scary to a teacher who has many students performing at about 67-77% mastery of the science content throughout the year. If these students' scores on the exam are about 13% lower because of the ELA difficulty, does that mean that this teacher can expect all students that
Literacy on Regents

normally perform at a C level to be expected to achieve less than 65% success and fail the exam? What about students that are achieving 65% mastery? Should a 90's student expect to get a 77 on the test? All these questions can question the validity of this test. The Regents Living Environment Exam tests students' content knowledge, but is also testing their ability to read questions and ELA skills as demonstrated in the data.

When the actual exam was re-worded, many difficult vocabulary words and sentence structures were adjusted to allow ease the impact of the reading. This adjustment allowed the science literacy to be more focused on the “questioning of evidence” (Roberts & Gott, 2004) rather than the questioning of reading ability. Students who struggle with longer words and sentences can get lost in the reading. If they can’t make it through the sentence, how can we expect them to pull the meaning of the content out of it? The biggest challenge that the accelerated students had mentioned was the struggle they had just figuring out what the question was asking. Once they knew what was being asked, they could easily answer the question. Lack of success for students on this test was due to getting lost in finding out what the question asked. Incorrect answers might have been from the “lack of knowledge or because of the inability to successfully comprehend the test and its items” (Homan, Hewitt, & Linder, 1994, p.349). Students simply did not know what the question was asking, even if they might have known the answer. It was interesting to see how this impacted the time it took for the students to complete the 30 questions. There was a 15 minute difference in time for competition between accelerated students that took test Version A and Version B. It is believed to be directly related to students knowing what the question asked. The accelerated classes are bright students who know that they need to find the meaning of the question before
answering it, and they have and use strategies they need to complete that task. The class that took the test with more difficult vocabulary and sentence structure was forced to spend more time finding the meaning of the questions. It was astonishing to see that two classes with equal ability had a 15 minute difference in time on a 30 question multiple choice test, not to mention the difference in success. Which strongly supports the belief that "the higher the item readability, the more the students miss that item" (Hewitt & Homan, 2004, p.1). When students were shown alternate the other version of the test that they did not take, reactions were similar across the board. All students overwhelmingly agreed that Version B "was so much easier." Students even complained about the difference in difficulty, "Why can't we take that test? Do you like them better?" There were even students, in the accelerated class that is composed of the top 25 students in the school, that when given the alternate test actually started to take the test not even realizing that it was the same. Smaller vocabulary words and shorter sentences make questions seem entirely different even when focusing on the exact same content. This only shows that a teacher needs to make sure they adjust their instruction to prepare the students to be able to answer many forms of the same question, having the students prepared to overcome the differences in literacy and readability of the test. Students who know the content also need to know the test and know the literacy level required. This makes it essential to analyze science programs and make sure testing is done at appropriate levels not only for content, but to make sure that the students are mature enough and skilled enough to handle the literacy skills required.

The results of the research have made it clear that English/Language Arts skills clearly are impacting student success on the exams that we are giving in classrooms.
Teachers must scramble to identify how their students are performing with both the content knowledge and in ELA skills. At times teachers may not realize how standardized their focus is and may “mistakenly feel that their assessment-oriented questions are sufficient to help struggling comprehension” (Fordham, 2006, p.394). Teachers need to be aware of this impact of the literacy on the exams and adjust their instruction accordingly to meet their students’ needs. “Ongoing opportunities for collaborative professional development” (Pedrotty-Bryant, Vaughn, Linan-Thompson, Ugel, Harniff, & Hougen, 2000, p.239) will allow for continued adjustment in professional practice which will allow for teachers to recognize and adjust what they are doing. Whether teachers seek professional development focused around becoming more of an ELA teacher within the science classrooms, or they search to find additional reading strategies to be used in their class, continued growth and education is needed for the teachers and administrators about how their students perform on tests because of literacy levels. There are very few research articles and resources available for teachers to use as a guide. Extensive research will provide few suggestions about improving literacy on science tests. Rather, it will be up to the educator to seek and find their own solutions that can be adjusted around their own students’ needs. Teachers who wish to seek solutions from other sources might find it more appropriate to find the solutions from their students. Assess what their students need to achieve success on exams, and determine what ELA skills are lacking and how they can impact success. Once teachers determine what is needed, they will need to trust their judgment as professionals and the judgment of their colleagues as to the way they can best suit student needs.
There was no noticeable difference in scores between genders when we look back at how individuals did overall. It was shocking to think that life science and biology are the areas of science that normally do not have the common stereotype about the boys being naturally stronger in science then the girls. The gender difference that is common in science doesn’t apply in this case since we are dealing with life science. Our data didn’t find any patterns that disagreed with the research stating that “boys excelled in Application in all the sciences except biology” (Tamir, 1988, p.143) and that “Although overall boys and girls who major in biology achieve equally well there are certain topics in which girls excel in Application, especially in topics related to human biology” (Tamir, p.134). There was no additional data among our students that supported or disproved that boy or girls are stronger in this area of science. It was interesting to first learn how there is normally no difference in gender compared to performance in life science, and it helps to explain why we also did not notice any pattern, granted no pattern was expected. No pattern in gender differences appeared, but there was a strong pattern that supports the mindset that 8th grade students will be greatly impacted in a negative way by the literacy requirements of the Living Environment Exam if they are all asked to take the test rather then just the accelerated students. Success rates are dependant on literacy above and beyond the content knowledge. How can students be expected to demonstrate content knowledge about a question or topic, if they don’t know or understand the question they are being asked? Even a student who knows everything won’t be successful if that student doesn’t know the question that they are being asked. The average 8th grade student will struggle with this level of literacy and have a less chance of being successful on the Regents Living Environment Exam in 8th grade.
**Recommendations for Future Research**

There are many areas where the research about literacy level can be taken in the future that were not addressed by this data. To find out when it is most appropriate for students to take this level of test so they can maximize their success and ability to demonstrate success, we would need to determine at what level the increased literacy level does not impact the students' ability to be successful. Rather then testing just the 8th grade students, future research should explore grades 7-12 to find out if there is a level where the increased literacy has little impact on expression of content knowledge. If we can determine this level, then we could administer this test and be assured that it is a valid test that does test just the students' knowledge and skills in the Living Environment content. It would also require us to not only research across grade levels, but also to research across topic areas, and see if this literacy impact is universal for all subjects and grades. It would also be handy to go beyond the topic area of science. All educators talk about how ELA skills impact student success for all areas of education, but there is never any number put with that statement. Similar tests should be given in Math, Social Studies, and ELA to determine what the percentages are and if there is any pattern or differences in the impact of literacy for the different areas and grades.

Future research also can easily add gender to an area of focus. Averages for classes were taken, there also could have been averages taken for males and females in each class to see if there is a gender gap at any level of science or any other subject matter. More could be determined as to why there is this gender gap? What is the cause and is there an obvious solution? Research has stated that there are differences between in performance based on gender at many different areas, but it would be nice to see how
much of a difference there is in our own classrooms. Apply this research to our
instruction, and improve or eliminate it impact.

The tests that were given only consisted of multiple choice questions as a means
to save time in a packed schedule due to curriculum requirements. It would be even more
valuable to see what kind of results and differences would stem from the students being
asked to take a complete Regents exam, during a 3 hour time slot. This could lead to
even wider differences in performance and success of the various groups of students. Not
only would students be asked to select prewritten answers, but to write their own. The
lack of the ability to express their own thoughts to explain their knowledge might prove
to be an additional weakness. Not only are students struggling with the ability to read
what the question is asking, but they then might struggle with what they are trying to say
in their answers. Parts B, C, and D of the Living Environment Exam all require student
constructed responses which don’t allow a chance for students to get questions right
based on a lucky guess. What impact does writing skills have on student success?

This test should also be given at the end of the year when all students have had
the opportunity to complete the course work and have received instruction on all the
content. Students in this data were at slight disadvantage because this test was given little
over half way through the year rather then at the end of the year. This actually makes it
difficult to determine what kind of impact ELA skills had on students success for certain
questions since we know the content knowledge was lacking to begin with.

Not much research is available about how to determine the reading level of high
stakes tests like that the students had taken. More research should be done to determine
what the actual reading levels various science tests are at to better help us understand
why literacy is having such a great impact on our students' success. If we can determine literacy level of the students, and literacy level of our test, then it is easier for us to determine at what grade level students can perform skills and demonstrate content knowledge to the best of their ability. We can also as professionals continue to more accurately adjust our instruction to better suit the needs of all our students in order to help they achieve success as often as possible. Re-wording difficult tests does increase a student’s ability to successfully demonstrate knowledge on that test. However, high stakes testing often doesn’t allow for re-wording to take place and a student will find themselves struggling not only to demonstrate content knowledge on a science test but literacy and ELA skills as well. A valid science test is meant to assess student science content knowledge and skills, lack of literacy skills can take away from that tests ability to do what it is supposed to do.
References


Technological Education, 22, 1, 5-21.


Appendix A
Test Version A

Part A
Answer all questions in this part. [30]

Directions (1-30): For each statement or question, write on your separate answer sheet the number of the word or expression that, of those given, best completes the statement or answers the question.

1. In the diagram below, what does X most likely represent?

- Consumers
- Producers
- Minerals in the soil
- (1) autotrophs
- (2) herbivores
- (3) decomposers
- (4) carnivores

2. Two closely related species of birds live in the same tree. Species A feeds on ants and termites, while species B feeds on caterpillars. The two species coexist successfully because
   (1) each occupies a different niche
   (2) they interbreed
   (3) they use different methods of reproduction
   (4) birds compete for food

3. After a hormone enters the bloodstream, it is transported throughout the body, but the hormone affects only certain cells. The reason only certain cells are affected is that the membranes of these cells have specific
   (1) receptors
   (2) tines
   (3) antibodies
   (4) carbohydrates

4. A characteristic of a DNA molecule that is not characteristic of a protein molecule is that the DNA molecule
   (1) can replicate itself
   (2) can be very large
   (3) is found in nuclei
   (4) is composed of subunits

5. The graph below illustrates the relative amounts of product formed by the action of an enzyme in a solution with a pH of 6 at seven different temperatures.

<table>
<thead>
<tr>
<th>Temperature (°C)</th>
<th>Amount of Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

Which statement best expresses the amount of product that will be formed at each temperature if the experiment is repeated at a pH of 4?
   (1) The amount of product formed will be equal to that produced at pH 6.
   (2) The amount of product formed will be greater than that produced at pH 6.
   (3) The amount of product formed will be less than that produced at pH 6.
   (4) The amount of product formed can not be accurately predicted.

6. Which statement best explains the fact that some identical twins appear different from one another?
   (1) Their DNA is essentially the same and the environment plays little or no role in the expression of their genes.
   (2) Their DNA is very different and the environment plays a significant role in the expression of their genes.
   (3) Their DNA is very different and the environment plays little or no role in the expression of their genes.
   (4) Their DNA is essentially the same and the environment plays a significant role in the expression of their genes.
7 Which statement best expresses the relationship between the three structures represented below?

Part of a protein molecule
Part of a DNA molecule

(1) DNA is produced from protein absorbed by the cell.
(2) Protein is composed of DNA that is produced in the cell.
(3) DNA controls the production of protein in the cell.
(4) Cells make DNA by digesting protein.

8 The diagram below represents a common laboratory technique in molecular genetics.

One common use of this technology is the
(1) production of a human embryo to aid women who are unable to have children
(2) change of single-celled organisms to multicellular organisms
(3) introduction of a toxic substance to kill bacterial cells
(4) production of hormones or enzymes to replace missing human body chemicals

9 Which statement provides accurate information about the technique illustrated below?

Dissected African cotton plant
Healthy American cotton plant

(1) This technique results in offspring that are genetically identical to the parents.
(2) New varieties of organisms can be developed by this technique known as selective breeding.
(3) This technique is used by farmers to eliminate mutations in future members of the species.
(4) Since the development of cloning, this technique is no longer used in agriculture.

10 Thousands of years ago, giraffes with short necks were common within giraffe populations. Nearly all giraffe populations today have long necks. This difference could be due to

(1) giraffes stretching their necks to keep their heads out of reach of predators
(2) giraffes stretching their necks so they could reach food higher in the trees
(3) a mutation in genetic material controlling neck size occurring in some skin cells of a giraffe
(4) a mutation in genetic material controlling neck size occurring in the reproductive cells of a giraffe

11 Estrogen has a direct effect on the

(1) formation of a zygote
(2) changes within the uterus
(3) movement of an egg toward the sperm
(4) development of a placenta within the ovary
12. A new chemical was discovered and introduced into a culture containing one species of bacteria. Within a day, most of the bacteria were dead, but a few remained alive. Which statement best explains why some of the bacteria survived?

(1) They had a genetic variation that gave them resistance to the chemical.
(2) They were exposed to the chemical long enough to develop a resistance to it.
(3) They mutated and became a different species after exposure to the chemical.
(4) They absorbed the chemical and broke it down in their digestive systems.

13. A current proposal in the field of classification divides life into three broad categories called domains. This idea is illustrated below.

Which concept is best supported by this diagram?

(1) Evolutionary pathways proceed only in one set direction over a short period of time.
(2) All evolutionary pathways will eventually lead to present-day organisms.
(3) All evolutionary pathways are the same length and they all lead to present-day organisms.
(4) Evolutionary pathways can proceed in several directions with only some pathways leading to present-day organisms.

14. After the union of sperm and egg, the single-celled zygote develops into a multicellular organism with specialized cells by the processes of

(1) meiosis and replication
(2) mitosis and differentiation
(3) cloning and growth
(4) fertilization and gamete production

15. A certain plant species, found only in one particular stream valley in the world, has a very shallow root system. An earthquake causes the stream to change its course so that the valley in which the plant species lives becomes very dry. As a result, the species dies out completely. The effect of this change on this plant species is known as

(1) evolution
(2) extinction
(3) mutation
(4) succession

16. When a planarian (a type of worm) is cut in half, each half usually grows back into a complete worm over time. This situation most closely resembles

(1) asexual reproduction in which a mutation has occurred
(2) sexual reproduction in which each half represents one parent
(3) asexual reproduction of a single-celled organism
(4) sexual reproduction of a single-celled organism

17. Which statement describes the reproductive system of a human male?

(1) It releases sperm that can be used only in external fertilization.
(2) It synthesizes progesterone that regulates sperm formation.
(3) It produces gametes that transport food for embryo formation.
(4) It shares some structures with the excretory system.

18. The immune system of humans may respond to chemicals on the surface of an invading organism by

(1) releasing hormones that break down these chemicals
(2) synthesizing antibodies that mark these organisms to be destroyed
(3) secreting antibiotics that attach to these organisms
(4) altering a DNA sequence in these organisms
19 Which statement about the gametes represented in the diagram below is correct?

(1) They are produced by females.
(2) They are fertilized in an ovary.
(3) They transport genetic material.
(4) They are produced by mitosis.

20 The dissolved carbon dioxide in a lake is used directly by

(1) autotrophs
(2) parasites
(3) fungi
(4) decomposers

21 Which transplant method would prevent the rejection of tissue after an organ transplant?

(1) using organs cloned from the cells of the patient
(2) using organs produced by genetic engineering to get rid of all proteins in the donated organs
(3) using organs only from pigs or monkeys
(4) using an organ donated by a close relative because the proteins will always be identical to those of the recipient

22 Ten breeding pairs of rabbits are introduced onto an island with no natural predators and a good supply of water and food. What will most likely happen to the rabbit population?

(1) It will remain relatively constant due to equal birth and death rates.
(2) It will die out due to an increase in the mutation rate.
(3) It will increase until it exceeds carrying capacity.
(4) It will decrease and then increase indefinitely.

23 Vaccinations help prepare the body to fight invasions of a specific pathogen by

(1) inhibiting antigen production
(2) stimulating antibody production
(3) inhibiting white blood cell production
(4) stimulating red blood cell production

24 All cells of an organism are engaged in many different chemical reactions. This fact is best supported by the presence in each cell of thousands of different kinds of

(1) enzymes
(2) nuclei
(3) chloroplasts
(4) organelles

25 Nutritional relationships between organisms are shown in the diagram below.

The mouse population would most likely decrease if there were

(1) an increase in the frog and tree populations
(2) a decrease in the snake and hawk populations
(3) an increase in the number of decomposers in the area
(4) a decrease in the amount of available sunlight
26 Even before a flower bud opens, certain plant chemicals have colored the flower in patterns particularly attractive to specific insects. At the same time, these chemicals protect the plant's reproductive structures by killing or inhibiting pathogens and insects that may feed on the plant. Which statement about the plant and the other organisms mentioned is correct?

(1) Chemicals affect plants but not animals.
(2) Organisms of every niche may be preyed on by herbivores.
(3) Any chemical produced in a plant can protect against insects.
(4) Organisms may interact with other organisms in both positive and negative ways.

27 A fire burns an oak forest down to bare ground. Over the next 150 years, if the climate remains constant, this area will most likely

(1) remain bare ground
(2) return to an oak forest
(3) become a rain forest
(4) become a wetland

28 Continued depletion of the ozone layer will most likely result in

(1) an increase in skin cancer among humans
(2) a decrease in atmospheric pollutants
(3) an increase in marine ecosystem stability
(4) a decrease in climatic changes

29 A change in the acidity of mountain lakes would most likely be a result of

(1) ecological succession of the area at the top of the mountain
(2) the introduction of new species into the lakes
(3) air pollution from smoke stacks miles away
(4) planting grasses and shrubs around the lakes

30 A forest is cut down and is replaced by a cornfield. A negative consequence of this practice is

(1) an increase in the carbon dioxide released into the atmosphere
(2) an increase in the size of predators
(3) a decrease in biodiversity
(4) a decrease in the amount of soil that is washed away during rainstorms
Appendix B

Test Version B

Regents Exam Practice Questions

**Directions:** For each statement or question, write on your separate answer sheet the number of the word or expression that, of those given, best completes the statement or answers the questions.

1. In the diagram, what type of living thing could the "X" be, that breaks down things to put minerals in the soil?

![Diagram](attachment:image.png)

1. autotroph  
2. herbivores  
3. decomposer  
4. carnivore

2. Two closely related bird species live in the same area, bird A eats ants and termites, while bird B eats caterpillars. They both can survive in that area because
   1. they have different niches, or do different things.
   2. they can interbreed
   3. they use different types of reproduction
   4. they compete against each other for the same food to eat

3. Hormones travel in the blood, and go to only certain types of cells. Each hormone only works on that certain cell because the outside of the cell has a specific _____________ that only matches that hormone.

1. receptor site  
2. tissue  
3. antibody  
4. carbohydrate

4. Which of the following is true about DNA molecule BUT NOT a protein molecule?

1. they can split to make more of themselves
2. they can be very complex
3. are in the nucleus
4. are made of smaller things.
The graph below shows how much stuff is made by an enzyme at different temperatures. The pH is 6 and does not change.

![Graph showing enzyme activity at different temperatures.]

5. Which statement best describes what the graph tells you about how the enzyme works at the pH of 4 with different temperatures?
   1. The amount of product is the same.
   2. The amount of product will be greater than at the pH of 6.
   3. The amount of product will be less than at the pH of 6.
   4. You don’t know anything about the pH of 4 from this graph.

6. Why can some identical twins can have different heights and weights even when they have identical DNA?
   1. Their DNA is the same, and their environment and personal decisions have no impact.
   2. Their DNA is different and the environment changes what their genes look like.
   3. Their DNA is different, and the environment and personal decisions have no effect.
   4. Their DNA is the same, but personal decisions can change what your body looks like.

7. Which statement is the most correct about cells, DNA, and protein in the diagram below?

![Diagram showing cell, DNA, and protein molecule.]

   1. DNA is made of protein that the cell absorbs.
   2. Protein is made of DNA produced in the cell.
   3. DNA controls how the cell makes protein in the ribosomes.
   4. Cells make DNA when they digest protein.
8. The diagram below shows a laboratory technique that puts human DNA into a bacteria cell.

How could this technique be helpful to humans?
1. produces embryos for women who can't have children
2. makes single celled organisms multicellular
3. kills bacteria with toxic stuff
4. creates bacteria that can produce a human hormones for people who can't make it.

9. Which statement is true based on the diagram below?
1. the offspring are identical to the parents
2. new varieties can be developed through selective breeding programs for farmers
3. farmers can prevent mutations in future species for farmers
4. cloning has made this technique no longer useful to farmers

10. Over thousands of years, giraffes necks went from being short to long. This is because
1. giraffes stretch their neck daily to keep it away from lions
2. giraffes stretch their necks daily to get food
3. a mutation for long necks randomly occurred in the DNA of the skin cells
4. a mutation of long necks randomly occurred in the DNA of reproduction cells
11. Estrogen causes the most changes in
   1. the zygote  2. uterus  3. egg & sperm movement  4. placenta development in the ovary.

12. A new chemical was put in a culture of bacteria. Almost all the bacteria were killed. Why might the few that lived survive?
   1. they had a genetic trait different from the others that was resistant to the chemical
   2. they were exposed long enough to it to learn to be resistant
   3. they changed their genes to become resistant
   4. they digested and broke down the chemical before it harmed them.

13. Which idea is supported by the diagram below?

```
Present      Bacteria  Archaea  Eukarya
          ↑
Past
```

   1. evolution occurs over a short period of time
   2. all evolution pathways become something today
   3. all evolution pathways are the same length of time and lead to a present day animal
   4. evolution pathways can be in different directions and may lead to something alive today

14. Which two processes are described as: a single cell dividing into many cells that become different?
   1. meiosis and replication
   2. mitosis and differentiation
   3. cloning and growth
   4. fertilization and gamete production

15. When a species completely dies out, it is called
   1. evolution  2. extinction  3. mutation  4. succession

16. When an organism grows back part of its body that is lost, this occurs:
   1. asexual reproduction with mutations
   2. sexual reproduction, half from each parent
   3. asexual reproduction in dividing cells
   4. sexual reproduction in dividing cells
17. Which is true about human male reproduction?
   1. releases sperm for fertilization outside of the bodies.
   2. progesterone regulates sperm production
   3. makes cells from food transportation
   4. shares some organs with the excretory system

18. When something bad gets in our body, we fight it by
   1. making hormones in the body to kill it
   2. make antibodies in the body to kill it
   3. make antibiotics in the body to kill it
   4. alter its DNA to kill it

19. Which statement about the sex cells below is true?

![Sex Cells]

   1. made by females
   2. fertilized in the ovary
   3. transport DNA
   4. produced by mitosis in body cells

20. Carbon dioxide in water is used for photosynthesis by
   1. autotrophs  2. parasites  3. fungi  4. decomposers

21. Which organs would be most likely accepted during a transplant and not recognized as foreign?
   1. identical organs cloned from the actual patient
   2. similar organs made from genetic engineering to get rid of all the proteins
   3. similar organs from monkey and pigs
   4. similar organs from a close relative

22. What will happen to the number of rabbits if nothing ever hunts them?
   1. number doesn’t change, birth and death rates are the same
   2. they all will die due to mutation rate increasing
   3. numbers will increase until the land can’t support them all
   4. numbers will decrease then increase forever

23. Vaccines help fight viruses by
   1. stopping foreign substances to be made
   2. starting antibody production
   3. stopping white blood cells from being made
   4. starting red blood cell production
24. Chemical reactions occur in all cells, to help speed up the reactions cells must have.
   1. enzymes
   2. nuclei
   3. chloroplasts
   4. organelles

25. The diagram below shows how food and energy is passed through some living things.

Which will cause the mouse population to DECREASE?
   1. increase in frogs and trees
   2. decrease in snakes and hawks
   3. increase in decomposers
   4. decrease in sunlight for the plants

26. Flowers have colored petals that attract insects for pollination. Some colors are caused by chemicals that kill insects and keep them from eating the flower. Which statement is true?
   1. chemicals affect plants and not animals
   2. organisms of every niche are eaten by herbivores
   3. all chemicals protect against insects
   4. organisms interact in both positive and negative ways

27. If a fire burns down a forest completely, in 150 years the area will most likely
   1. stay bare
   2. return to being a forest
   3. become a rain forest
   4. become a wetland

28. Continuing to destroy the ozone will
   1. increase skin cancer from the sun
   2. decrease atmosphere pollution
   3. increase in ocean stability
   4. decrease in weather changes
29. Increasing acid in a mountain’s lake is most likely caused by
   1. increased ecological succession
   2. new species in the lake
   3. air pollution from smoke stacks
   4. planting trees and shrubs around lakes

30. Why would cutting a forest down and replacing it with a corn field be bad?
   1. it increases the carbon dioxide released
   2. it increases the size of the predators
   3. it decreases the different kinds of species living in the area
   4. it decreases the amount of soil eroded away